

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:04:33 ; Search time 137.122 Seconds
(without alignments)
3768.837 Million cell updates/sec

Title: US-09-836-602-6

Perfect score: 1509

Sequence: 1 MPRGTALCLLSLLPSGFS.....LVTSGALLAVLIGTGYFLMN 301

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlp

-Q/cgn2_1/USPTO_spool/US09836602/runat_29102002_101143_3758/app_query.fasta_1.1429
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602@cgn_1.1.312@runat_29102002_101143_3758 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB Length	ID	Description
1	1509	100.0	906	24	AAD22663	Human CD34 surface
2	1509	100.0	951	24	AAD22662	Human CD34 surface
3	1509	100.0	1122	24	AAD22661	Human full-length
4	1509	100.0	2463	22	AAH75132	Nucleotide sequenc
5	1509	100.0	2615	21	AAF20898	Human CD-34 polynu
6	1509	100.0	2615	21	AAA34776	Human adenosine re
7	1509	100.0	3490	21	AAF20899	Human CD-34 polynu
8	1509	100.0	3490	21	AAA34777	Human adenosine re
9	839	55.6	1065	18	AAT59508	Porcine CD34 from
10	508	33.7	615	21	AAC00249	Human secreted pro
11	311	20.6	233	22	ABA51134	Human breast cell
12	311	20.6	233	22	ABA69132	Human foetal liver
13	311	20.6	233	22	ABA36064	Probe #14530 for g
14	311	20.6	233	22	AAK17442	Human brain expres
15	311	20.6	233	22	AAK43242	Human bone marrow
16	311	20.6	233	22	AAI24018	Probe #13951 for g
17	311	20.6	233	22	AAI49317	Probe #18003 used
18	311	20.6	233	22	AAI09602	Probe #9593 used t
19	265	17.6	209	22	ABA50740	Human breast cell
20	265	17.6	209	22	ABA68710	Human foetal liver
21	265	17.6	209	22	ABA35674	Probe #14140 for g
22	265	17.6	209	22	AAK17050	Human brain expres
23	265	17.6	209	22	AAK42833	Human bone marrow
24	265	17.6	209	22	AAI23597	Probe #13530 for g
25	265	17.6	209	22	AAI48910	Probe #17596 used
26	265	17.6	209	22	AAI09212	Probe #9203 used t
27	233	15.4	235	22	AAS39236	Novel human diagno
28	170.5	11.3	406	22	ABA45618	Human breast cell
29	170.5	11.3	406	22	ABA56124	Human foetal liver
30	170.5	11.3	406	22	ABA25775	Probe #4241 for ge
31	170.5	11.3	406	22	AAK04311	Human brain expres
32	170.5	11.3	406	22	AAK29806	Human bone marrow
33	170.5	11.3	406	22	AAI14396	Probe #4329 for ge
34	170.5	11.3	406	22	AAI35770	Probe #4456 used t
35	170.5	11.3	406	22	AAI04219	Probe #4210 used t
36	159	10.5	13154	20	AAI13275	Enterococcus faeca
37	150.5	10.0	2336	23	ABL25662	Drosophila melanog
38	148	9.8	40875	18	AAT80043	Insert from cosmid
39	147.5	9.8	7720	21	AAAS3800	Genomic DNA encodi
40	145.5	9.6	4116	23	ABL24389	Drosophila melanog
41	145.5	9.6	5162	23	ABL20278	Drosophila melanog
42	145.5	9.6	6116	23	ABL24388	Drosophila melanog
43	144.5	9.6	8298	22	AAK72613	Human immune/haema
44	144	9.5	5163	19	AAV20700	Cryptosporidium pa
45	144	9.5	5163	21	AAA61849	ORF encoding a por

ALIGNMENTS

RESULT 1

AAD22663
ID AAD22663 standard; DNA; 906 BP.

XX AAD22663;

XX 26-FEB-2002 (first entry)

XX Human CD34 surface antigen deleted variant (dCD34) encoding DNA.

XX Human: surface marker; surface antigen; T lymphocyte; gene therapy;
KW CD34 deleted variant; dCD34; ds.

XX Homo sapiens.

XX Synthetic.

XX Key

XX Location/Qualifiers

FT CDS 1..906

FT /*tag= a
 FT /product= "Human CD34 surface antigen
 FT deleted variant (dCD34)"
 XX

PN EP1148066-A1.

XX 24-OCT-2001.

XX 18-APR-2001; 2001EP-0109374.

XX 18-APR-2000; 2000DE-1019075.

XX (ZAND/) ZANDER A R.

XX Zander AR;

XX WPT; 2002-019289/03.

DR P-PSDB; AAE13543.

XX New gene transfer vector (accession number DSM13396) containing a
 PT transgene and a nucleic acid sequence coding for a surface marker.
 PT useful in gene therapy, and for detecting genetically modified cells or
 PT cells which do not express CD34 -

XX Claim 3; Page 14-15; 28pp; English.

XX The patent discloses a gene transfer vector (accession number DSM13396)
 CC containing a transgene and a nucleic acid sequence coding for a surface
 CC marker. The surface marker is the CD34 surface antigen, its fragment or
 CC variant. The vector is useful for in vitro transduction of T lymphocytes,
 CC for gene therapy, and in the enrichment, detection and analysis of cells
 CC in vitro that do not naturally express CD34. T lymphocytes transduced
 CC with the vector are also useful in gene therapy. The CD34 nucleic acid
 CC sequences (marker genes), their fragments or variants are used for
 CC detecting genetically modified cells or cells which do not naturally
 CC express CD34. The present sequence is a DNA encoding human CD34 surface
 CC antigen deleted variant (dCD34).

XX Sequence 906 BP; 247 A; 250 C; 199 G; 210 T; 0 other;

XX Alignment Scores:

Pred. No.: 2.47e-110 Length: 906
 Score: 1509.00 Matches: 301
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0

US-09-836-602-6 (1-301) x AAD22663 (1-906)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 1 ATGCGGGGGCTGGACCGCGCTTTGCTGCTGAGTTGCTGCGCTTCTGGGTTTCATGAGT 60
 Qy 21 LeuAspAsnAnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 61 CTTGACAAACACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTCACAAAGTT 120
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 121 TCTACAATGTATCTCTACCAAGAACTACACACCTAGTACCTTGGAAAGTACCAAGCTG 180
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValIys 80
 Db 181 CACCCTGTGCTCAACATGGCAATGAGCCACAAACATACAGAAACGACAGTCAAA 240
 Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 241 TTCACATCTACCTCTGTGATAACCTCAGTTATGGAACACAAACTCTTCTGTCCAGTCA 300
 Qy 101 GlnThrSerValIleSerThrValPheThrThrThrProAlaAsnValSerThrProGluThr 120
 Db 301 CAGACCTCTGTGTAATCAGCAGAGTGTTCACCAACCCCAACGTTTCAACTCCAGAGACA 360

Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 Db 361 ACCTTGAAGCCTACCTGCTACCTGGAAATGTTTCAGACCTTTCAACACCTAGCAGTACG 420
 Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
 Db 421 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTCTATCCTTAAGTGACATCAAG 480
 Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 Db 481 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCTGCCTGGAG 540
 Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 Db 541 CAAAATAGACCTCCAGCTGTGCGGAGTTTAAAGAGACAGGGGAGAGGGCTGGCCCGA 600
 Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 Db 601 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGCTGGGGCCAGGTATGCTCCTGCTC 660
 Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
 Db 661 CTTGCCAGTCTGAGTGTGAGGCTCAGTGTCTACTGTGTGCTTGGCCAAACAGAACAA 720
 Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
 Db 721 ATTTCCAGCAACTCCCACTTATGAAAAGACCAACTGTGACCTGAAAAGCTGGGATC 780
 Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
 Db 781 CTAGATTTCACTAGCAGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAGACCTGATT 840
 Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
 Db 841 GCATGTGTCACTCGGAGGCCCTGCTGCTGTCTTGGGCATCACTGGCTATTCTCTGATG 900
 Qy 301 Asn 301
 Db 901 AAT 903
 RESULT 2
 AAD22662
 ID AAD22662 standard; DNA; 951 BP.
 XX
 AC AAD22662;
 XX
 DT 26-FEB-2002 (first entry)
 XX
 DE Human CD34 surface antigen truncated variant (tCD34) encoding DNA.
 KW Human; surface marker; surface antigen; T lymphocyte; gene therapy;
 KW CD34 truncated variant; tCD34; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..951
 FT /*tag= a
 FT /product= "Human CD34 surface antigen
 FT truncated variant (tCD34)"
 FT
 PN EP1148066-A1.
 PD 24-OCT-2001.
 XX
 PF 18-APR-2001; 2001EP-0109374.
 XX
 PR 18-APR-2000; 2000DE-1019075.
 XX
 PA (ZAND/) ZANDER A R.
 XX

PI Zander AR;
XX WPI; 2002-019289/03.
DR P-PSDB; AAE13542.
XX
PT New gene transfer vector (accession number DSM13396) containing a
PT transgene and a nucleic acid sequence coding for a surface marker,
PT useful in gene therapy, and for detecting genetically modified cells or
PT cells which do not express CD34 -
XX
PS Claim 3; Page 12-13; 28pp; English.
XX
CC The patent discloses a gene transfer vector (accession number DSM13396)
CC containing a transgene and a nucleic acid sequence coding for a surface
CC marker. The surface marker is the CD34 surface antigen, its fragment or
CC variant. The vector is useful for in vitro transduction of T lymphocytes,
CC for gene therapy, and in the enrichment, detection and analysis of cells
CC in vitro that do not naturally express CD34. T lymphocytes transduced
CC with the vector are also useful in gene therapy. The CD34 nucleic acid
CC sequences (marker genes), their fragments or variants are used for
CC detecting genetically modified cells or cells which do not naturally
CC express CD34. The present sequence is a DNA encoding human CD34 surface
CC antigen truncated variant (tCD34).
XX
SQ Sequence 951 BP; 261 A; 264 C; 213 G; 213 T; 0 other;

Alignment Scores:
Pred. No.: 2,63e-110 Length: 951
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-09-836-602-6 (1-301) x AAD22662 (1-951)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
DB 1 ATGCCGCGGGCTGGACCGCGCTTTGCTGCTGAGTTTGTGCTGCTTCTGGGTTTCATGAGT 60
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 61 CTTGACAAACAGCGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 120
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 121 TCTCAATGTAATCTACCAAGAACTACACACCTAGTACCTCTGGAGTAGTACAGCCCTG 180
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
DB 181 CACCCTGTGTCCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA 240
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 241 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTCTCTCCAGTCA 300
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 301 CAGACCTCTGTAAATCAGACAGTGTTCACCCAGCCAGCAAGCTTCACTCCAGAGACA 360
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
DB 361 ACCTTGAAGCCCTAGCCTGTGCTACCTGGAAATGTTTTCAGACCTTTCAACCACTAGCAGT 420
QY 141 LeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIleLys 160
DB 421 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTTCTCTATCTTCTAGTGCATCAAG 480
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
DB 481 GCAGAAATCAATGTTTACGATCATCAGAAAGTGAATGACTCAGGCGATCTGCTGGAG 540
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluLeuAlaArg 200

Db 541 CAAATAAGACCTCCAGCTGTGCGAGTTTAAGAAGGACAGGAGAGGGCTGCGCCGA 600
QY 201 ValLeuCysGlyGluGluAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
DB 601 CTGCTGTGTGGGAGGAGCAGCTGATGCTGATGCTGGGGCCAGGTATGCTTCCCTGCTC 660
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
DB 661 CTTGCCAGTCTGAGGTGAGGCTCAGTGTCTACTGCTGCTTGGCCCAACACAGANA 720
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLysLysLeuGlyIle 260
DB 721 ATTTCCAGCAAACTCCAATTATGAAAAAGCACCAATCTGACCTGAAAAAGCTGGGATC 780
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
DB 781 CTAGATTTTCACTGACCAAGATGTTGCAAGCCACACAGAGCTATTTCCAAAAGACCTGATT 840
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
DB 841 GCACCTGGTCACTCGGAGCCCTGCTGCTGCTTGGGCATCACTGGCTATTTCTCTGATG 900
QY 301 Asn 301
DB 901 AAT 903

RESULT 3
AAD22661
ID AAD22661 standard; DNA; 1122 BP.
XX
AC AAD22661;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human full-length CD34 (fICD34) surface antigen encoding DNA.
XX
KW Human; surface marker; surface antigen; T lymphocyte; gene therapy;
KW full-length CD34; fICD34; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 1..1122 /*tag= a
FT /product= "Human full-length CD34 surface antigen"
XX
PN EP1148066-A1.
XX
PD 24-OCT-2001.
XX
PF 18-APR-2001; 2001EP-0109374.
XX
PR 18-APR-2000; 2000DE-1019075.
XX
PA (ZAND/) ZANDER A R.
XX
PI Zander AR;
XX
DR WPI; 2002-019289/03.
DR P-PSDB; AAE13541.
XX
PT New gene transfer vector (accession number DSM13396) containing a
PT transgene and a nucleic acid sequence coding for a surface marker,
PT useful in gene therapy, and for detecting genetically modified cells or
PT cells which do not express CD34 -
XX
PS Claim 3; Page 9-11; 28pp; English.
XX
CC The patent discloses a gene transfer vector (accession number DSM13396)
CC containing a transgene and a nucleic acid sequence coding for a surface
CC marker. The surface marker is the CD34 surface antigen, its fragment or
CC variant. The vector is useful for in vitro transduction of T lymphocytes,

CC for gene therapy, and in the enrichment, detection and analysis of cells
 CC in vitro that do not naturally express CD34. T lymphocytes transduced
 CC with the vector are also useful in gene therapy. The CD34 nucleic acid
 CC sequences (marker genes), their fragments or variants are used for
 CC detecting genetically modified cells or cells which do not naturally
 CC express CD34. The present sequence is a DNA encoding human full-length
 CC CD34 (f1CD34) surface antigen.
 XX

SQ Sequence 1122 BP; 305 A; 311 C; 270 G; 236 T; 0 other;

Alignment Scores:

Pred. No.: 3,25e-110 Length: 1122
 Score: 1509.00 Matches: 301
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0

US-09-836-602-6 (1-301) x AAD22661 (1-1122)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 DB 1 ATGCGGGGGCTGGAGCCGGCTTTGCTGTGAGTTGCTGCTTCTGGGTTTCATGAGT 60
 QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 DB 61 CTTCACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 120
 QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 DB 121 TCTACAATGTATCTACCAAGAAACTACAAACACTAGTACCCTTGGGAAGTACCAGCGTG 180
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
 DB 181 CACCCTGTGTTCAACTATGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA 240
 QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 DB 241 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACAACAACACTCTCTGTCAGTCA 300
 QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 DB 301 CAGACCTCTCTAATCAGCAGAGTGTTCACCACCCAGCCCAACGTTTCAACTCCAGAGACA 360
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 DB 361 ACCTTTGAAGCCTAGCCGTGTCACCTGGAAATGTTTCAGACCTTTCAACCATAGCAGTACG 420
 QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
 DB 421 CTTCACAACATCTCCCACTAAACCTATACATCATCTCTCTATCTAAGTGACATCAAG 480
 QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 DB 481 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATGACTCAGGGCATCTGCTGGAG 540
 QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 DB 541 CAAAATGAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGGGGAGGCGCTGGGCCGA 600
 QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlnValCysSerLeuLeu 220
 DB 601 GTCTGTGTGGGAGGAGGAGGCTGATGCTGATGCTGGGGCCCGAGGTATGCTCCCTGCTC 660
 QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
 DB 661 CTTCGCCAGTCTCAGGTGAGGCGCTCAGTGTCTACTGCTGTCTTGGCCAAACAGAACAGAA 720
 QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLysLysLeuGlyIle 260
 DB 721 ATTTTCAGCAAACTCAACTTATGAAGAACCACTCACTGACCTGAAAGAGCTGGGGATC 780
 QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280

DB 781 CTAGATTTCACCTGAGCAAGATGTTGCAAGCCACGACAGCTATTCCCAAGACCCCTGATT 840
 QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
 DB 841 GCACCTGTGTCACCTCGGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCCTGATG 900
 QY 301 Asn 301
 DB 901 AAT 903
 RESULT 4
 ID AAH75132 standard; DNA; 2463 BP.
 AC AAH75132;
 XX 13-NOV-2001 (first entry)
 DE Nucleotide sequence of a human CD34 polypeptide.
 KW Human; CD34 gene; blast crisis; chronic myelogenous leukemia;
 XX nm23-H4 kinase gene; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 91..1248
 FT /*tag= a
 FT /product= "CD34"
 XX WO200164946-A1.
 XX 07-SEP-2001.
 XX 28-FEB-2001; 2001WO-JP01485.
 XX 02-MAR-2000; 2000JP-0058043.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Mano H, Miyazato A, Ueno S, Yoshida K, Yamanaka T, Ikeda U;
 PI Shimada K, Hatake K, Ozawa K, Asada K, Kato I;
 XX WPI: 2001-550191/61.
 DR P-PSDB; AAG67120.
 XX Method for detecting chronic myelogenous leukemia by comparing
 PT expression levels of CD34 and nm23-H4 genes -
 XX Disclosure; Page 36-40; 60pp; Japanese.
 CC The present sequence encodes a human CD34 polypeptide. The
 CC specification describes a method of detecting blast crisis in chronic
 CC myelogenous leukemia. The method comprises comparing the amounts of
 CC expression of at least two genes in a sample, particularly CD34 gene
 CC and nm23-H4 kinase gene. The method allows the worsening stages of
 CC chronic myelogenous leukemia to be easily detected at a high
 CC reliability.
 XX
 SQ Sequence 2463 BP; 589 A; 709 C; 580 G; 585 T; 0 other;
 Alignment Scores:
 Pred. No.: 9e-110 Length: 2463
 Score: 1509.00 Matches: 301
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0
 US-09-836-602-6 (1-301) x AAH75132 (1-2463)
 QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20

Db 127 ATGCCCGGGGTGGACCGGCTTTGCTGCTGAGTTTCTGCTTCTGGTTCATGACT 186
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACAAACAGGGTACTCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 246
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTACAAATGTATCTTACCAGAAACTACACACCTAGTACCTTGGAAAGTACCAGCCGT 306
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 307 CACCCTGTCTCAACATGCAATGAGGCCACCAACAAACATCACAGAAACGACAGTCAA 366
QY 81 PheThrSerThrSerValIleThrSerValTyrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATACCTCAGTTTATGGAACACAAACTCTTCTGTCTCCAGTCA 426
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTAATCAGCACAGTGTTCACCCAGCCCAACGTTTCACTCCAGAGACA 486
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 140
Db 487 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTCAGACCTTTCACACCTAGCAGTACG 546
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 547 CTTCGCAACATCTCCCACTAAACCTATACATCATCTCTCCATCTTAAGTGACATCAAG 606
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 607 GCAGAAATCAATGTTTCAGGCATCAGAGAAGTCAATTCACCTCAGGGCATCTGCCTGGAG 666
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 667 CAAAATAAGACCTCCACGCTGTGCGGAGTTTAAAGAGCAGAGGGGAGGCGCTGGCCCGA 726
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 727 GTGCTGTGTGGGAGAGCAGGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCTC 786
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 787 CTTGCCAGTCTCAGGTGAGGCTCAGTCTACTGCTGCTGCTGTGCTGTGSCCAACAGACAGAA 846
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 847 ATTTCCAGCAACTCCAACTTATGAAAAAGCACCACCAATCTGACCTGAAAAAGCTGGGGATC 906
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 907 CTGATTTTCACTGAGCAAGATGTTGCAACCCAGCAGAGTATTCCTCCAAAGACCCCTGATT 966
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 967 GCACCTGTCACCTCGGGAGCCCTGCTGGTGTCTTGGGCATCAGTGGCTATTCTCTGATG 1026
QY 301 Asn 301
Db 1027 AAT 1029

RESULT 5

AAF20898

ID AAF20898 standard; DNA; 2615 BP.

XX AAF20898;

XX 14-MAR-2001 (first entry)

XX Human CD-34 polynucleotide fragment #2465.

XX

KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cycostatic;
KW respiratory obstruction; pulmonary vasoconstriction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
XX Homo sapiens.
XX WO2000062736-A2.
XX
XX 26-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-US08020.
XX
XX 06-APR-1999; 99US-0127958.
XX
XX (UYEC-) UNIV EAST CAROLINA.
XX (NYCE/) NYCE J W.
XX
XX Nyce JW;
XX
XX WPI; 2000-679539/66.
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
XX trigger adenosine receptors during metabolism, useful e.g. for treating
XX cancers and respiratory obstructions -
XX
XX Disclosure; Page 292-293; 1592pp; English.

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'Universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors and immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAF18434 to AAF21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:

Pred. No.:	9.72e-110	Length:	2615
Score:	1509.00	Matches:	301
Percent Similarity:	100.00%	Conservatives:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-836-602-6 (1-301) x AAF20898 (1-2615)

Qy 1 MetProArgGlyTTPThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 294 ATGCCGGGGCTGACCGCGCTTTCGTCTGAGTTGCTGCGCTTCGGGTTTCATGAGT 353
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 354 CTTGACACACAGGTACTGCTACCCAGAGTACCTACCCAGGGAGACATTTTCAATGTT 413
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 414 TCTACAAATGTATCTACCAAGAACTACAACACCTAGTACCTTGGAAATACAGCCTG 473
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 474 CACCTGTGTCTCAACATGGCAATGAGGCCACAACAACATCACAGAAACACAGTCAA 533
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 534 TTCACATCTACTCTGTGATAACTCAGTTATGGAACACACAAACTCTTCGTCCAGTCA 593
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 594 CAGACCTCTGTAATCAGCACAGTGTTCACCACCCAGCCCAACGTTTCAACTCCAGAGACA 653
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 654 ACCTTGAAGCCTAGCCTGTACCTGGAATGTTTCAGACCTTTCAACACACAGCAGTACG 713
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
Db 714 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTATCTCAAGTGACATCAAG 773
Qy 161 AlaGluLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 774 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCTGCTGGAG 833
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 834 CAAATATAGACCTCCAGCTGTGCGGAGTTTAAGAGGACAGGGAGAGGCGCTGGCCGA 893
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 894 GTGCTGTGGGGAGGAGCAGGCTGATGCTGCTGGGGCCAGGATGCTCCCTGCTC 953
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 954 CTTGCCAGTCTGAGGTGAGGCCCTCAGTGTCTACTGTGCTTTGGCCACACAGAACAGAA 1013
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGlyLys 260
Db 1014 ATTTCCAGCAACATCCCACTTATGAAGAGCACCACATCTGACCTGAAGAGCTGGGATC 1073
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1074 CTAGATTTCTAGTACGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAGAACCCCTGATT 1133
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 1134 GCATGTGTACCTCGGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCTCTGATG 1193
Qy 301 Asn 301
Db 1194 AAT 1196

RESULT 6
AAA34776
ID AAA34776 standard; DNA: 2615 BP.
XX
AC
AC
XX

28-JUL-2000 (first entry)
Human adenosine receptor related polynucleotide SEQ ID NO:2465.
Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
Homo sapiens.
WO200009525-A2.
24-FEB-2000.
03-AUG-1999; 99WO-US17712.
03-AUG-1998; 98US-0095212.
(UYEC-) UNIV EAST CAROLINA.
Nyce JW;
WPI; 2000-205971/18.
New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -
Disclosure; Page 602; 1343pp; English.
The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airflow, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.
SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:
Pred. No.: 9.72e-110 Length: 2615
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-6 (1-301) x AAA34776 (1-2615)

```
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 294 ATCCGGGGGGCTGGACCGGCTTGGCTGCTGAGTTTGGCTTCTGGTTTCATGAGT 353
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 354 CTTGACAAACAGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 413
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 414 TCACAAATGTATCCTACCAAGAACTACAAACACCTAGTAGTCCCTTGGAGTACCAGCTG 473
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 474 CACCCTGTGCTCAACATGCCAATGAGGCCACCAACAACATCAGAAACAGACAGTCAA 533
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 534 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 593
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 594 CAGACCTCTGTATCAGCACAGTGTCCACCACCCCAACGCTTCAACTCCAGAGACA 653
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 654 ACCTTGAAGCCTAGCCTGTCCACCTGGAATGTTTCAGACCTTTCACACCTAGCAGTACG 713
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 714 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTCCTAACTGACATCAAG 773
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 774 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAAATTTGACTCAGGGGCATCTGCCTGGAG 833
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyLeuAlaArg 200
Db 834 CAATAATGACCTCCACCTGCTGGAGTTTAAAGNAGCAGGGGAGGCGCTGGCCCCGA 893
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlnValCysSerLeuLeu 220
Db 894 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCTC 953
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 954 CTTGCCCAGTCTGAGGTGAGGCGCTCAGTGTCTACTGTGTCTTGGCCAAACAGAACAGAA 1013
QY 241 IleSerSerLysLeuGlnLeuMetLysHisGlnSerAspLeuLysLysLeuGlyLe 260
Db 1014 ATTTCCAGCAAACTCCAACTTATGAAAACACCAACATCTGACCTGAAAAGCTGGGGATC 1073
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1074 CTAGATTTCACTGAGCAAGATGTTGCAAGCCACCAGAGCTATTCCTCCAAAAGACCTGATT 1133
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 1134 GCACGTGCTACCTGGGAGGCGCTGCTGGCTGCTTGGGCATCACTGGCTATTTCCTGATG 1193
QY 301 Asn 301
Db 1194 AAT 1196
RESULT 7
AAF20899
ID AAF20899 standard; DNA: 3490 BP.
XX AAF20899;
AC AAF20899;
XX
DT 14-MAR-2001 (first entry)
XX
```

Human CD-34 polynucleotide fragment #2466.

Low adenosine antisense oligonucleotide; phosphorothioate; allergy; human; airway disorder; bronchoconstriction; lung inflammation; surfactant depletion; respiratory; bronchodilator; antiinflammatory; immunosuppressive; antiasthmatic; analgesic; hypotensive; cytotstatic; respiratory obstruction; pulmonary obstruction; impeded respiration; surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS; respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis; pulmonary hypertension; emphysema; pulmonary transplantation rejection; chronic obstructive pulmonary disease; pulmonary infection; bronchitis; cancer; ss.

Homo sapiens.

WO200062736-A2.

26-OCT-2000.

24-MAR-2000; 2000WO-US08020.

06-APR-1999; 99US-0127958.

(UYEC-) UNIV EAST CAROLINA.

(NYCE/) NYCE J W.

Nyce JW;

WPI; 2000-679539/66.

Low adenosine (A) content antisense oligonucleotides which do not trigger adenosine receptors during metabolism, useful e.g. for treating cancers and respiratory obstructions -

Disclosure; Page 291-292; 1592pp; English.

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytotstatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and or activity of target polypeptides associated with the lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAF18434 to AAF21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;

Alignment Scores:

Pred. No.:	1.41e-109	Length:	3490
Score:	1509.00	Matches:	301
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0
US-09-836-602-6 (1-301) x AAF20899 (1-3490)			
Qy	1	MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	1169	ATGCCGGGGCGGTGAGCCGCCTTTCCTGCTGAGTTTGCTGCTTCTGGGTTTCATGAGT	1228
Qy	21	LeuAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	1229	CTTGACAAACGGTACTGCTACCCAGAGTTACCTACCACGGGAACATTTCCAATGTT	1288
Qy	41	SerThrAsnValSeryrGlnGluThrThrProSerThrLeuGlySerThrSerLeu	60
Db	1289	TCTACAATGTATCCTACCAAGAATACTACACACCTTAGTACCCTTGGAGTACCAGCCTG	1348
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThraSnIleThrGluThrThrValLys	80
Db	1349	CACCCTGTCTCAACATGCAATGAGGACCAACAAACATCACAGAACGACAGTCAA	1408
Qy	81	PheThrSerThrSerValIleThrSerValTyrrGlyAsnThrAsnSerSerValGlnSer	100
Db	1409	TTACATCTACCTCTGTGATPAACCTCAGTTTATGGAAACACAAACTCTCTGTGCCAGTCA	1468
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	1469	CAGACCTCTGTAATCAGCAGCAGTGTCCACCACCCAGCCAACCTTCAACTCCAGAGACA	1528
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer	140
Db	1529	ACCTTGAAGCCTAGCCTGTCCACCTGGAATTTTTCAGACTTTCACACCTAGCCTAGC	1588
Qy	141	LeuAlaThrSerProThrLysProTyrrThrSerSerProIleLeuSerAspIleLys	160
Db	1589	CTTGCAACATCTCCCCTAACCCTATACATCATCTCTCTATCTAAGTGACATCAAG	1648
Qy	161	AlaGluIleLysCysSerGlyIleArGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	1649	GCAGAAATCAAAATGTTTCAGGCATCAGAGAAAGTGAATTCAGTTCAGGCGCATCTGCCTGGAG	1708
Qy	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	1709	CAAAATAAGACCTCCAGCTGTGCGGAGTTTAAGAAGACAGGGGAGGGCCTGGCCGA	1768
Qy	201	ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	1769	GTGCTGTGCGGAGGAGCAGGCTGATGCTGATGCTGGGCCCCAGGTATGCTCCCTGCCTC	1828
Qy	221	LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
Db	1829	CTTGCCCCAGTCTGAGGTGAGGCTCAGTGTCTACTGCTGTGCTTGGCCCAAGACAGAA	1888
Qy	241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle	260
Db	1889	ATTTCCAGCAAACTCCAACTTATGAAAAAGCACCAATCTGACCTGAAAAGCTGGGGATC	1948
Qy	261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrrSerGlnLysThrLeuIle	280
Db	1949	CTAGATTTTCACTGAGCAAGATGTTCAGCCACCAGAGCTATTTCCTCAAAAGACCCGTGAT	2008
Qy	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrrPheLeuMet	300
Db	2009	GCACCTGCTACCTCGGAGCCCTGCTGGCTGCTTGGGCATCACTGGCTATTTCCTGATG	2068
Qy	301	Asn	301
Db	2069	AAT	2071
RESULT 8			
AAA34777			
ID	X AAA34777 standard; DNA; 3490 BP.		

```
US-09-836-602-6 (1-301) x AAA34777 (1-3490)
Qy 1 MetProArgGlyThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1169 ATGCCGGGGCTGGACGGCTTGTGCTGAGTTGCTGCGCTTCTGGGTTTCATGAGT 1238
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 1229 CTTGACAAACAGGTACTGCTACCCAGAGGTACTACCCAGGGAACATTTTCAATGTT 1288
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 1289 TCTCAAAATGTATCTACCAAGAACTACAACTAGTACCTTGAAGTAGTACCAGCGCTG 1348
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 1349 CACCTCTGTCTCAACATGGCAATGAGGCCACAAACAACATCACAGAAACGACAGTCAA 1408
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 1409 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACAACAACTCTTCTGTCCAGTCA 1468
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 1469 CAGACCTCTGTAAATCAGCACAGTGTTCACCACCCAGCAACGTTTCAACTCCAGAGACA 1528
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 1529 ACCTTGAAGCTTACCTGTACCTGGAATGTTTTCAGACCTTTTCAACCACTTAGCAGCTAGC 1588
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
Db 1589 CTTGCAACATCTCCCATTAACCTATACATCACTTCTCTATCTCCTAAGTGACATCAAG 1648
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 1649 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCCTGGAG 1708
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 1709 CAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAGACAGGGGAGAGGGCTGGCCCGA 1768
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 1769 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATCTCTGGGCCCAAGGTATGCTCCTGCTC 1828
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 1829 CTTGCCAGTCTGAGTGAGGCTCAGTGTCTACTGTGCTTGGTCTTGGCCAAACAGAACAA 1888
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 1889 ATTTCCAGCAACCTCAACTATGAAAGACCAACCAATCTGACCTGAAAGAGCTGGGGATC 1948
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1949 CTAGATTCTACTGAGCAAGATGTTGCAAGCCACCAAGACTATTTCCCAAGAACCTTGATT 2008
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 2009 GCACCTGGTCACTCGGGAGCCCTGCTGGCTGCTTGGGCATCACTGGCTATTTCCTGATG 2068
Qy 301 Asn 301
Db 2069 AAT 2071
RESULT 9
AAT59508
ID AAT59508 standard; cDNA; 1065 BP.
XX
AC AAT59508;
XX
```

```
DT 27-OCT-1997 (first entry)
XX Porcine CD34 from clone 1AX2-3 encoding cDNA.
DE Antibody; bone marrow chimera; graft; antigen; pig; immune response;
KW haematopoietic progenitor cell; stem cell; cord blood; ss.
XX
XX Sus scrofa.
XX Key Location/Qualifiers
PH CDS 31..1065
FT /*tag= a
FT /*product= CD34
FT sig_peptide 31..123
FT /*tag= b
FT mat_peptide 124..1062
FT /*tag= c
XX
XX WO9640244-A1.
XX 19-DEC-1996.
XX 03-JUN-1996; 96WO-US08340.
XX 07-JUN-1995; 95US-0475634.
XX (BIOT-) BIOTRANSPLANT INC.
XX Hawley RJ, Monroy RL;
PI WPI; 1997-108628/10.
DR P-PSDB; AAW11822.
XX
XX Antibodies selective for CD34 + porcine cells - used to isolate such
XX cells for generating bone marrow chimera(s) that will accept porcine
XX grafts, also new porcine CD34 antigen
XX Example 1; Fig 1; 35pp; English.
XX
XX The present sequence encodes porcine CD34. Antibodies (Ab) have been
XX produced that recognise porcine CD34+ cells. The Ab's are used to
XX recover CD34+ cells, preferably from porcine bone marrow but also from
XX cord blood, and CD34+ cells, which are enriched in haematopoietic
XX progenitor cells, especially stem cells, are used to generate bone
XX marrow chimeras in a human to assist acceptance of a porcine graft
XX (i.e. to prevent or inhibit an immune response against the graft).
XX Ab's can also be used to detect CD34+ cells, e.g. to determine if
XX mixed chimerism has been established. The CD34 polypeptide and its
XX fragments are used to raise Ab.
XX
XX Sequence 1065 BP; 248 A; 325 C; 273 G; 219 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 2,5e-57 Length: 1065
Score: 839.00 Matches: 184
Percent Similarity: 68.63% Conservative: 37
Best Local Similarity: 57.14% Mismatches: 75
Query Match: 55.60% Indels: 26
DB: 18 Gaps: 6
US-09-836-602-6 (1-301) x AAT59508 (1-1065)
Qy 1 MetProArgGlyThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 67 ATGCCGGGGTGGACCGCTCTGCTTGTGAGTTGCTGCGCTTCTGGGTTTCATGAGT 126
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuPro----- 32
Db 127 GTG---AACAGCTCAACTATTGCTTCCACCTGGCAGCTGGCGTGGGTCAACTCCCACC 183
Qy 33 -----ThrGlnGlyThrPheSerAsn----- 39
Db 184 GGGCGGGCTACCGCAGGGGCGAGCTATCATCCGGGTCAACTATCTCAGACATATCTTCACT 243
```

```
QY 40 ValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSer 59
D 244 GTTCTTACAATATATCCAAACAGGAGAACACATCA--GATGCTTTCGAAAGTGCCAGC 300
QY 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVal 79
D 301 CTCACACTGTCTCTCAGGGCAGCAGTGGGACCAACCGCATCTCAGCCCTACAGTT 360
QY 80 LysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGln 99
D 361 AATTTCATGCTACCTCGCGGTCACCTCGTCCCGGAAACCGTTAACTCTCTGTCACG 420
QY 100 SerGlnThrSerValIleThrSerValPheThrThrProAlaAsnValSerThrProGlu 119
D 421 CCTCAGACCTCT--CTAGCCACAGGCTCTCCGCCACCATCACTTTACAACTTCAGAG 477
QY 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThr 139
D 478 GTGACCTCTCAGCCCGCAGCAGCTTCCCGAGGAATGTTTCAGACCCCTCTACAAACAGTACC 537
QY 140 SerLeuAlaThrSerProThrLysProTyrThrThrSerSerSerProfileLeuSerAspIle 159
D 538 AGCCTGCGAGATCCCCACAGCCCTACACATCATCTCTCCCTACCCAGGTAGCCAC 597
QY 160 LysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
D 598 AAGGGGGAAGTCAAAATGTGCCAAATCAAGAGGTGAATTGACCAAGGTATCTGCCGTG 657
QY 180 GluGlnAspLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
D 658 GAGCAATGAGACCTCCCGCTCGGAGAGTTTAAAGAGGCAATGGAGAGAGTGTATG 717
QY 200 ArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
D 718 CAATCTGTGTGGCAGAGCAGGCTGAGCGCGGCCAGGG-----GTGTGCTCTGTG 771
QY 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThr 239
D 772 CTCCTTGCCCAATCTGAGTGAACCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831
QY 240 GluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGly 259
D 832 GAACCTAGCAGCAAGTTCCTGCTTCTGGAAGACCAAGTCTGAACCTGAGAGATGAGC 891
QY 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu 279
D 892 ATCCAAACTTCTCGAACAAGATGTAGGACCCACAGAGACTCTCCGAAAGACCTTG 951
QY 280 IleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeu 299
D 952 ATTGCACCTGGTCACCTCGGGGATCCTGCTGGCTGTCTTGGGCATCCTGCTACTTGCTG 1011
QY 300 MetAsn 301
D 1012 ATGAAC 1017
RESULT 10
AAC00249
ID AAC00249 standard; cDNA: 615 BP.
XX
AC AAC00249;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 247.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX
KW gene therapy; chromosome mapping; ss.
OS Homo sapiens.
PN EP1033401-A2.
```

```
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG00243.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS Claim 1; SEQ ID 247; 7lpp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 615 BP; 149 A; 181 C; 159 G; 119 T; 7 other;
```

```
Alignment Scores:
Pred. No.: 1,7e-31 Length: 615
Score: 508.00 Matches: 99
Percent Similarity: 96.12% Conservative: 0
Best Local Similarity: 96.12% Mismatches: 4
Query Match: 33.66% Indels: 0
DB: 21 Gaps: 0
```

US-09-836-602-6 (1-301) x AAC00249 (1-615)

```
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
D 305 ATGCGCGGGGCTGGACCGCTTTGCTGCTGAGTTTCTGCTTCTGGTTTCATGAGT 364
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
D 365 CTTGACAAACACGGTACTGCTACCCAGAGTTTACCTACCCAGGAAACATTTTCAAATGTT 424
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
D 425 TCTCAAAATGATCTCTACCAAGAAACTACACACCTGTGACCTTGGAGAGTACCACCTG 484
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
D 485 CACCCTGTGCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTSSNG 544
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
D 545 TTCACATCTTASCTGTGTGTATAAASCTCAGTTTATGGRACACAAACTCTTCTGTCCAGTCA 604
QY 101 GlnThrSer 103
D 605 CAGACCTCT 613
DB
RESULT 11
ABA51134/c
```

ID ABA51134 standard; DNA; 233 BP.
 XX
 AC ABA51134;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #9829.
 XX
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 4; SEQ ID NO 9829; 327bp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and Br 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
 Pred. No.: 1.76e-16 Length: 233
 Score: 311.00 Matches: 60
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.61% Indels: 0
 DB: 22 Gaps: 0

US-09-836-602-6 (1-301) x ABA51134 (1-233)

Qy 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
 Db TCTGGGTTTCATGAGTCTTGACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGA 172

Qy 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu 55
 Db 171 ACATTTCAATGTTTACAAATGTATCCTACCAGAACTACAACTAGTACCTT 112
 Qy 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
 Db 111 GGAAGTACCAGCTGCACCTGTGTCTCAACATGCAATGAGGCCACAAACAATCACA 52

RESULT 12
 ABA69132/c

ID ABA69132 standard; DNA; 233 BP.

XX

AC ABA69132;

XX 01-FEB-2002 (first entry)

DT Human foetal liver single exon nucleic acid probe #17437.

DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

OS WO200157277-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -

XX Claim 4; SEQ ID NO 17437; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
 Pred. No.: 1.76e-16 Length: 233
 Score: 311.00 Matches: 60
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.61% Indels: 0
 DB: 22 Gaps: 0

US-09-836-602-6 (1-301) x ABA69132 (1-233)

Qy 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
 Db TCTGGGTTTCATGAGTCTTGACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGA 172

US-09-836-602-6 (1-301) x AAK17442 (1-233)

Qy	16	SerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGly	35
Db	231	TCTGGGTTCATGAGCTTGACAAACAGGCTACTGCTACCCAGAGTTACTACCCAGGGA	172
Qy	36	ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu	55
Db	171	ACATTTTCAAAATGTTCTCAAAATGTACTCAAGAAACTACAACACTAGTACCCCTT	112
Qy	56	GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr	75
Db	111	GGAAGTACCAAGCTGCACCTGTGTCTCAACATGGCAATGAGGCCACAACAACATCACA	52

RESULT 15

AAK43242/C

ID AAK43242 standard; DNA; 233 BP.

AAK43242:

06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 17799.

Human; bone marrow expressed exon; gene expression analysis; probe; microarray; cancer; leukaemia; lymphoma; mveloma; ss.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000: 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.
yy

XX
PA (MOI-E-) MOLECULAR DYNAMICS INC

XX PI Penn SG. Hanzel DK. Chen W. Rank DR.

XX
DP WPT: 2001-188900/53

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human bone marrow.

XX
PS Example 4: SEO ID NO: 17799: 658pp + Sequence Listing: English

AA The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.

Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other:

Alignment Scores:

Pred. No.:	1.76e-16	Length:	233
Score:	311.00	Matches:	60
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	20.61%	Indels:	0
DB:	22	Gaps:	0

US-09-836-602-6 (1-301) x AAK43242 (1-233)

Qy	16	SerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGly	35
Db	231	TCTGGGTTCATGAGTCTTGACAAACACGGTACTGCTACCCAGAGTTACCTACCCAGGGA	172
Qy	36	ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu	55
Db	171	ACATTTTCAATGTTCTCAAAATGTATCTACCAAGAAACTACAACACCTAGTACCCCTT	112
Qy	56	GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr	75
Db	111	GGAAGTACCAGCTGCACCTGTGTCTCAACATGGCAATGAGGCCACACAACATCATCA	52

Search completed: October 30, 2002, 08:25:52

Job time : 143.122 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 30, 2002, 08:08:13 : Search time 38 6131 Seconds
(without alignments)
1914.781 Million cell updates/sec

Title: US-09-836-602-6
Perfect score: 1509
Sequence: 1 MPRGWTALCLSLPSGFMS.....LVTSGALLVLGITGVFLMN 301

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlip
-Q=/cgn2_1/USPTO.spool/US09836602/runat_29102002_101145_3837/app_query.fasta_1.1429
-DB=Issued_Patents_NA -Qfmt=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602@cgn1_1_44@runat_29102002_101145_3837 -NCPU=6 -ICPU=3
-WARN_TIMEOUT=30 -THREADS=1 -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	839	55.6	1065	2	US-08-475-634D-18
2	144	9.5	5163	3	US-08-700-651-1
3	144	9.5	5163	3	US-08-928-361B-4
4	144	9.5	5318	3	US-08-700-651-2
5	144	9.5	5318	3	US-08-928-361B-3
6	137.5	9.1	5511	3	US-08-928-361B-2
7	137.5	9.1	7334	3	US-08-928-361B-1
8	129	8.5	1505	1	US-07-915-246-1
9	128	8.5	390	4	US-09-197-649-7
10	127.5	8.4	3337	1	US-08-072-610-1
11	127.5	8.4	3337	2	US-08-719-822B-1
12	127.5	8.4	3337	4	US-09-092-458-1
					Sequence 18, Appl
					Sequence 1, Appl
					Sequence 4, Appl
					Sequence 2, Appl
					Sequence 3, Appl
					Sequence 2, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 7, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 1, Appl

13	125.5	8.3	2589	1	US-08-325-367A-3	Sequence 3, Appl
14	125.5	8.3	2685	3	US-08-362-525-21	Sequence 21, Appl
15	125.5	8.3	4614	1	US-08-325-267A-1	Sequence 1, Appl
16	123	8.2	2754	1	US-08-270-076A-10	Sequence 10, Appl
17	121	8.0	2214	6	5258502-1	Patent No. 5258502
18	121	8.0	3168	4	US-09-165-239A-3	Sequence 3, Appl
19	119.5	7.9	688	4	US-08-998-416-915	Sequence 915, App
20	119.5	7.9	1107	2	US-08-991-300-1	Sequence 1, Appl
21	116	7.7	2793	1	US-08-209-747-1	Sequence 1, Appl
22	116	7.7	2793	1	US-08-458-298-1	Sequence 1, Appl
23	115.5	7.7	2032	4	US-09-241-581B-5	Sequence 5, Appl
24	115.5	7.7	2032	5	PCT-US95-07721-5	Sequence 5, Appl
25	115	7.6	2584	3	US-08-758-662-8	Sequence 8, Appl
26	114.5	7.6	2093	1	US-08-287-001A-1	Sequence 1, Appl
27	114.5	7.6	2093	5	PCT-US95-00941-1	Sequence 1, Appl
28	113	7.5	36519	3	US-08-923-137-2	Sequence 2, Appl
29	110.5	7.3	2164	4	US-08-760-615-3	Sequence 3, Appl
30	108.5	7.2	4108	4	US-08-981-729-8	Sequence 8, Appl
31	108.5	7.2	4108	4	US-08-981-446B-1	Sequence 1, Appl
32	108	7.2	2150	2	US-08-861-464-13	Sequence 13, Appl
33	108	7.2	2150	2	US-08-396-001-13	Sequence 13, Appl
34	108	7.2	2150	4	US-09-323-433A-13	Sequence 13, Appl
35	108	7.2	9636	1	US-08-323-170B-1	Sequence 1, Appl
36	108	7.2	9636	4	US-08-954-441-1	Sequence 1, Appl
37	107.5	7.1	3833	1	US-08-917-320-18	Sequence 18, Appl
38	107.5	7.1	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
39	107.5	7.1	5931	3	US-08-783-774-1	Sequence 1, Appl
40	106	7.0	10803	3	US-09-080-044-1	Sequence 1, Appl
41	106	7.0	19227	3	US-09-090-793-13	Sequence 13, Appl
42	106	7.0	40138	3	US-09-090-793-12	Sequence 12, Appl
43	105	7.0	3141	2	US-08-658-665-66	Sequence 66, Appl
44	105	7.0	3141	4	US-08-796-101-30	Sequence 30, Appl
45	105	7.0	3141	4	US-09-085-273-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-08-475-634D-18
; Sequence 18, Application US/08475634D
; Patent No. 5962644
; GENERAL INFORMATION:
; APPLICANT: HAWLEY, Robert J.
; APPLICANT: MONROY, Rodney L.
; TITLE OF INVENTION: Antibodies to Porcine CD34 Positive
; TITLE OF INVENTION: and Uses Therefor
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS 6.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,634D
; FILING DATE: June 7, 1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 61750-141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 973-994-1700
; TELEFAX: 973-994-1744
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 1065 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-475-634D-18

Alignment Scores:
Pred. No.: 1,65e-81 Length: 1065
Score: 839.00 Matches: 184
Percent Similarity: 68.63% Conservative: 37
Best Local Similarity: 57.14% Mismatches: 75
Query Match: 55.60% Indels: 26
DB: 2 Gaps: 6

US-09-836-602-6 (1-301) x US-08-475-634D-18 (1-1065)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 67 ATCCGGGGGCTGGACCGCTCTGCTGCTGAGTTGCTGCCCTCTGGTTTCACAGCT 126
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuPro----- 32
Db 127 GTG---AACAGCTCAACTATTGCTTCCACCTTCCAGCTGCCGCTGGGTCAACTCCACC 183
QY 33 -----ThrGlnGlyThrPheSerAsn----- 39
Db 184 GGCGCGGCTACCCGAGGGGAGGTATCATCCGGGTCAACTATCTCAGACATATCTTCCACCT 243
QY 40 ValSerThrAsnValSerTyGlnGluThrThrProSerThrLeuGlySerThrSer 59
Db 244 GTTCTTACAAATATATCCACGAGGAAACACACATCA--GATGCTTTCGAAAGTCCAGC 300
QY 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVal 79
Db 301 CTCACACTGTCTCTCAGGSCAGCAGTGGGACCCAGTATGCCATCTCAGGCCCTCAGGTT 360
QY 80 LysPheThrSerThrSerValIleThrSerValTyGlyAsnThrAsnSerSerValGln 99
Db 361 AATTTTCATGCTACCTCGGGGTCACCCCTCGTCCCGAACCCTTAACCTCTCTGTCCAG 420
QY 100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
Db 421 CCTCAGACCTCT---CTAGCCACAGCGTCTCTCGCCACCATCACTTACAACTTCAGAG 477
QY 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 139
Db 478 GTACCCCTGCAGCCGACGACGCTTCCCGAGGAATGTTTCAGACCCCTCTACAAACAGTACC 537
QY 140 SerLeuAlaThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIle 159
Db 538 AGCCCTGCGAGATCCCCACCCACCCCTACACATCATCTCTCTCTACCCAGGTAGCCAC 597
QY 160 LysAlaGluLeuLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Db 598 AAGGGGGAAGTCAAAATGTGCCAAATCAAAGAGGTGAATGACCAAGGTATCTGCCCTG 657
QY 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Db 658 GAGCGAAATGAGACCTCCGGCTCGGAGAGTTTAAAGAGCAATGGAGAGATTCATG 717
QY 200 ArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
Db 718 CAATCTGTGTGGGAGGAGAGGTGAGGCTGAGCGCGGCCAGGG-----GTGTGCTCCTTG 771
QY 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThr 239
Db 772 CTCCTTGGCCCAATCTGAGGTGAACCTCTACTGCTGCTGCTGTGCTGTGGCAACGGAACA 831
QY 240 GluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGly 259
Db 832 GAACCTAGCAGCAAGTTCCTGCTTCTGGAAAGACCCAGTCTGAACCTGAGAGAGATGAGC 891

; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5163
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-1

Alignment Scores:
Pred. No.: 5,74e-05 Length: 5163
Score: 144.00 Matches: 37
Percent Similarity: 47.37% Conservative: 26
Best Local Similarity: 27.82% Mismatches: 69
Query Match: 9.54% Indels: 1
DB: 3 Gaps: 0

US-09-836-602-6 (1-301) x US-08-700-651-1 (1-5163)

QY 22 AspAsnAsnGly-ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 41
Db 613 GACACAAACAACAACAACGACGACAAACAACAACAACAACAACAACAACAACAACA 672
QY 41 rThrAsnValSerTyGlnGluThrThrThrProSerThrLeuGlySerThrSerLeuH 61
Db 673 GACACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 732
QY 61 sProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLysPh 81
Db 733 AACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 792
QY 81 eThrSerThrSerValIleThrSerValTyGlyAsnThrAsnSerSerValGlnSerG 101
Db 793 TACACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 852
QY 101 nThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 121
Db 853 AACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 912
QY 121 rLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrThrSerLe 141
Db 913 CAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 972
QY 141 uAlaThrSerProThrLysProTyThrThrSerSerSer 153
```


Qy 204 GlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGln 223
Db 3334 ATCGAAAGATGACGATGATTGAAATGGAATGCATTTTACAATGATTCCAAATGATGAC 3393
Qy 224 SerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGluLeu----- 241
Db 3394 ACGCATGTCGCTTCAGATTTTAAAGTTTAAAGATGATGGGACACTATTTCAGTAAGATGC 3453
Qy 242 -----SerSerLysLeuGlnLeuMetLysLysHisGlnSerLysLeuLysLysLeu 258
Db 3454 GGAAGAGTGCAGTAACACTCGATTCAGATAGAGT----- 3492
Qy 259 GlyTleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThr 278
Db 3493 -----TTGGATTTCACAATTCCCTCCAGTAGCTGCGCATACACGCTGTCA----- 3537
Qy 279 LeuTleAlaLeuValThrSerGlyAlaLeuLeuAlaVal 291
Db 3538 ATAATAGTTGGTGTGAGCGCGGTGGAAAAATTCACGTA 3576
RESULT 8
US-09-836-602-6 (1-301) x US-07-915-246-1 (1-1505)
; Sequence 1, Application US/07915246
; Patent No. 5401836
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Chris L.
; APPLICANT: Fallis, Lynne
; APPLICANT: Bellmare, Guy
; APPLICANT: Bolvin, Rodolphe
; TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
; TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held, and Malloy
; STREET: 500 W. Madison St. 34th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/915,246
; FILING DATE: 19920716
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 91 P 1125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 707-8889
; TELEFAX: 312 707-9155
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1505 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; STRAIN: Westar
; DEVELOPMENTAL STAGE: Somatic
; TISSUE TYPE: Root
US-07-915-246-1

Alignment Scores: 0.000289 Length: 1505
Pred. No.:

Score: 129.00 Matches: 47
Percent Similarity: 48.78% Conservative: 33
Best Local Similarity: 28.66% Mismatches: 70
Query Match: 8.55% Indels: 14
DB: 1 Gaps: 5
US-09-836-602-6 (1-301) x US-07-915-246-1 (1-1505)
Qy 12 SerLeuLeuProSerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeu 31
Db 1370 AGTGCATACCTCCCTCCCTCTCTCCGCCACCTCCAGCTCCACCCACCGCTCTCTCC 1311
Qy 32 ProThrGlnGlyThrPheSer-----AsnValSerThrAsnValSerTyr 46
Db 1310 ACCACCGTGTCTCTCTCTCCACCCACCCACCGTATCCCTCTCCAGCTCCACCTCCGCGC 1251
Qy 47 GlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSerGlnHis 66
Db 1250 ACCATAACCCACCGTGTGAGCTCCACCGCGGAGAACCTCTCTCCACCGCTCTCTTT 1191
Qy 67 GlyAsnGluAlaThrThrAsnIle-----ThrGluThrValLysPheThrSer 83
Db 1190 TCCCCCGCTCCACCCACCTCCATGTCCACCTGCACCGCCACCTCCATATCTCTCCAGC 1131
Qy 84 ThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGlnThrSer 103
Db 1130 ACCAGCGCTTCTCCACCCACCGTATCCACCCGCTGTGCACCTCCACCTCTCTCCACCGCC 1071
Qy 104 ValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrThrLeuLys 123
Db 1070 TCCATTTCCACCGCTCCACCCAC-----TCCATGTCCACCTGCACCGCCACCC- 1023
Qy 124 ProSerLeuSerProGlyAsnValSerAspLeuSerThrThr-----SerThrSerLeuAla 142
Db 1022 ---TCCATATCTCTCTCCACCGACCGACCTCTCTCCACCCACCGTATCCACCATCGCGC 966
Qy 143 ThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLysAlaGlu 162
Db 965 ACCGCCAGCTCTCCACCCACCCACCTCTCTCGCTTCTCTCCACCTCCACCATGTCTCTCCAGC 906
Qy 163 IleLysCysSer 166
Db 905 TCCTCCGCTCTCC 894
RESULT 9
US-09-197-649-7
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Pribnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/C1-CON
; CURRENT APPLICATION NUMBER: US/09/197,649
; CURRENT FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: 07/829,461
; EARLIER FILING DATE: 1992-01-31
; EARLIER APPLICATION NUMBER: 07/739,055
; EARLIER FILING DATE: 1991-08-01
; EARLIER APPLICATION NUMBER: 07/561,968
; EARLIER FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.

RESULT 12
US-09-092-458-i/c
; Sequence 1, Application US/09092458
; Patent No. 6231861
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,458

CONSENT REFLECTION DATA:
APPLICATION NUMBER: US/09/092,458

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/719,821

```
,
, FILING DATE: 09/30/96
, ATTORNEY/AGENT INFORMATION:
,
, NAME: Cogoris, Adda
, REGISTRATION NUMBER: 29,714
, REFERENCE/DOCKET NUMBER: 5986/
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (212)527-7700
, TELEFAX: (212)753-6237
, TELEX: 236687
, INFORMATION FOR SEQ ID NO: 1:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 337 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: double
, TOPOLOGY: linear
, MOLECULE TYPE: DNA (genomic)
, HYPOTHETICAL: NO
, ANTI-SENSE: NO
, ORIGINAL SOURCE:
, ORGANISM: Plasmodium vivax
, IMMEDIATE SOURCE:
, CLONE: pMB3.3.1
US-09-092-458-1
```

Alignment Scores:	
Pred. No.:	3337
Score:	0.00166
Length:	127.50
Matches:	53
Percent Similarity:	41.88%
Best Local Similarity:	27
Query Match:	27.75%
Mismatches:	96
Indels:	15

```
DB: 4 Gaps: 6
US-09-836-602-6 (1-301) x US-09-092-458-1 (1-3337)
QY 5 TrpThrAlaLeuCysLeuLeuSerLeuLeuPro-----SerGlyPheMetSerLeu 21
DB 2527 TGGTAGTCATCGATGATAGATAAAATCCAGAGCCCTAAATCGTCTTCTTTGTTG 2468
QY 22 AspAsn-----AsnGlyThrAlaThrProGluLeuProThrGlnGlyThr 36
DB 2467 TCTAATGTATATCTCTCCGATGGTATTAATAATGTAATACATCTTCTCCCTCTCT 2408
QY 37 PheSerAsnValSerThrAsnValSerTyrglnGlnThrThrProSerThrLeuGly 56
DB 2407 GGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2348
QY 57 SerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGlu 76
DB 2347 GGTACTTCTACTCTACTACCGCTGGTACCTCTACTCTCTCTACTCTCTCTCTCT 2288
QY 77 ThrValThrValLysPheThrSerThrValIleThrSerValTyrglyAsnThrAsnSer 96
DB 2287 ACTTCTCTGGTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2228
QY 97 SerValGlnSerGlnThrSerValIleSerThrValPheThrProAlaAsnValSer 116
DB 2227 TCCACTTCTCTGGTACCTCT-----TCCACTTCTCTACTCTCTCGGGGTACCTCT 2174
QY 117 ThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThr 136
DB 2173 ACTTCTCTGGTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2120
QY 137 ThrSerThrSerLeuAlaThrSerProThr---LysProTyThrSerSerSerProIle 155
DB 2119 TCTTCTACTCTCGAGGTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2063
QY 156 LeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGln 175
DB 2062 ---TCTACGGTTCTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2006
QY 176 GlyIleCysLeuGluGlnAsnLysThrSerSer 186
DB 2005 GGAGTTGCCTCTAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1973
RESULT 13
US-08-325-267A-3
; Sequence 3, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, YERKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; NUMBER OF INVENTIONS: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2589 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: ABXL-ID
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2589
; US-08-325-267A-3
Alignment Scores:
Pred. No.: 0.00177 Length: 2589
Score: 125.50 Matches: 72
Percent Similarity: 44.03% Conservative: 46
Best Local Similarity: 26.87% Mismatches: 90
Query Match: 8.32% Indels: 61
DB: 1 Gaps: 12
US-09-836-602-6 (1-301) x US-08-325-267A-3 (1-2589)
QY 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
DB 1099 ACAACACGACACTGAA---CCATGGACTGGTACTTTTACTTCGACTTCCACTGAAATGTCT 1155
QY 46 -----TyrGlnGluThr-----ThrPro 52
DB 1156 ACTGTCACTGGAAACCAATGGCTTGCACACTGTATGATGTTGTGTCATAAACTCCA 1215
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
DB 1216 ACTACTGCCATCTCATCCAGTTTGCA-----TCATCATCTTCAGGACAAATCACCAGC 1269
QY 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValTyrgly 92
DB 1270 TCTATCAGC-----TCTTCGCGTCCCAATATTATACCCCATCTCTATCCT 1311
QY 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
DB 1312 AGCAATGGAACTCTGTGATTTCT---TCCTCAGTAATTTCT----- 1350
QY 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
DB 1351 TCCTCAGTCACCTCTCTCTATTCATCTCTCTCCAGTCATTTCTCTCCTCAGTCATTTCT 1410
QY 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
DB 1411 TCTTCTACAAACAACCTCCACTTCTATATTTCTGAATCATCTAAATCATCCGCTCATTTCCA 1470
QY 146 ThrLysProTyThrSerSerSerProIleLeuSerAspIleLysAlaGluIleLysCys 165
DB 1471 ACCAGTAGTTCACCTCTGGTTCT-----TCTGAGAGCGAAGACG 1509
QY 166 SerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSer 185
```

```

;
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: pY1105
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2685
; OTHER INFORMATION: /product= "Flocculation protein" /gene= nFLO1"
; US-08-362-525-21
Alignment Scores:
Pred. No.: 0.00188 Length: 2685
Score: 125.50 Matches: 72
Percent Similarity: 44.03% Conservative: 46
Best Local Similarity: 26.87% Mismatches: 90
Query Match: 8.32% Indels: 61
DB: 3 Gaps: 12

US-09-836-602-6 (1-301) x US-08-362-525-21 (1-2685)
Qy 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
Db 1234 ACAACCACCACTGAA---CCATGGACTGGTACTTTTACTTCGACTTCCTGAAATGCT 1290
Qy 46 -----TyrGlnGluThrThr-----ThrPro 52
Db 1291 ACTGTCACCTGGAACCAATGGCTGCCAACTGATGAAACTGTCAATGTTGTCAAAACTCCA 1350
Qy 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
Db 1351 ACTATGCCATCTCATCCAGTTTGTC-----TCATCATCTTCAGGACAAATCACCAGC 1404
Qy 73 AsnIleThrGluThrValLysPheThrSerThrValIleThrSerValThrGly 92
Db 1405 TCTATCAG-----TCTTCGCGTCCAATATTATCCCATCTTCTATCTCT 1446
Qy 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
Db 1447 AGCAATGGAACCTTCTGTGATTTCT---TCTCATGTAATTTCT----- 1485
Qy 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
Db 1486 TCCTCAGTCACCTTCTCTCTATTCACCTTCTTCCAGTCATTTCTCTCCTCAGTCATTTCT 1545
Qy 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
Db 1546 TCTTCTACAAACACCTCCACTTCTATATTTTCTGAATCATCTAAATCATCCGTCATTTCCA 1605
Qy 146 ThrLysProThrThrSerSerSerProIleLeuSerAspIleLysAlaGluIleLysCys 165
Db 1606 ACCAGTAGTCCACCTCTGGTTCT-----TCTGAGCGGAACG 1644
Qy 166 SerGlyIleArgGluValLysLeuThrGlnGlyLeuGluGlnAsnLysThrSer 185
Db 1645 AGTTCAGCTGGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1704
Qy 186 -----SerCysAlaGluPheLysLysAspArgGly 195
Db 1705 ACATATTCTTCTTTCATCATTTACCACCTTGTATTACAGTCGACACACAGCCAGGAATGCT 1764
Qy 196 GluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAla----- 210
Db 1765 TCTTCATTAC--CACCTGCTACCACTACAAAAACGACGACCAACCACTTGTGTTACCG 1822
Qy 211 ---AspAlaGlyAlaGlnValCysSerLeuLeuAlaGlnSerGluValArgProGln 229
Db 1823 TGACATCCTCGGAGTCTCATGTCGTGTCACCTGAATCCATCTCCCTCCGATGTTGTTCCACG 1882
Qy 230 CysLeuLeuValLeuAlaAsnArgThrGluIle---SerSerLysLeuGlnLeuMet 248
;
; LENGTH: 2685 base pairs
; TYPE: nucleic acid

```

RESULT 14

```

US-08-362-525-21
; Sequence 21, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLIS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARB & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KORULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 213289/T7020(V)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2685 base pairs
; TYPE: nucleic acid

```

Db 1883 CTACTGTTACTGTTACGGCGTCACACAGAGTATACACATGGTGCCTATTCTTACTACTA 1942

QY 249 LysLysHisGlnSerAspLeuLys 256
Db 1943 CAGAGACACAAAGCAACCAAG 1966

RESULT 15

US-08-325-267A-1
; Sequence 1, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO.
; APPLICANT: PENNTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; TITLE OF INVENTION: CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: ABXL-1D
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4614
US-08-325-267A-1

Alignment Scores:
Pred. No.: 0.00479 Length: 4614
Score: 125.50 Matches: 72
Percent Similarity: 44.03% Conservative: 46
Best Local Similarity: 26.87% Mismatches: 90
Query Match: 8.32% Indels: 61
DB: 1 Gaps: 12

US-09-836-602-6 (1-301) x US-08-325-267A-1 (1-4614)
QY 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
Db 3124 ACACACCACCTGAA---CCATGGACTGGTACTTTTACTTCGACTTCCACTGAAATGTCT 3180
QY 46 -----TyrGlnGluThrThr-----ThrPro 52
Db 3181 ACTGTCACTGGAACCAATGGCTTGCCCAACTGATGAAACTCTCATTTGTCAAAACACTCCA 3240
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
Db 3241 ACTACTGCCATCTCATCCAGTTGTCA-----TCATCATCTTTCAGGACAAACACACAG 3294
QY 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValTyrGly 92
Db 3295 TCTATCACG-----TCTTCGCGTCCCAATTATTACCCCAATTCATCTATCCT 3336
QY 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
Db 3337 AGCAATGGAACTCTCTGATTCT---TCTCAGTAATTCT----- 3375
QY 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
Db 3376 TCCTCAGTCACTTCTTCTCTATTCACTTCTTCTCCAGTCATTTCTTCTCCTCAGTCATTCT 3435
QY 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
Db 3436 TCTTCTACAACAACCTCCCACTTCTATATTTCTGAATCATCTAAATCATCCGTCATTTCCA 3495
QY 146 ThrLysProTyrThrSerSerSerProIleLeuSerAspIleLysAlaGluIleLysCys 165
Db 3496 ACCAGTAGTTCACCTCTGGTTCT-----TCTGAGACCGAAACG 3534
QY 166 SerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSer 185
Db 3535 AGTTCAGCTGGTTCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3594
QY 186 -----SerCysAlaGluPhelLysLysAspArgGly 195
Db 3595 ACATATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3654
QY 196 GluGlyLeuAlaArgValLeuLysCysGlyGluGluGlnAlaAspAla----- 210
Db 3655 TCTTCTATTAC--CACCTGTCTACCACTACAAAACAGCGGACAAACACCACTTTGGTTACCG 3712
QY 211 ---AspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGlnSerGluValArgProGln 229
Db 3713 TGACATCCTCGGAGTCTCATGTGTGCACATCCATCCCTCCCTCGGATTTGTTCCACAG 3772
QY 230 CysLeuLeuLeuValLeuAlaAsnArgThrGluIle---SerSerLysLeuGlnLeuMet 248
Db 3773 CTACTGTTACTGTTAGCGGCTCACACAGAGTATACACATGGTGCCTTATTTCTACTA 3832
QY 249 LysLysHisGlnSerAspLeuLys 256
Db 3833 CAGAGACACAAAGCAACCAAG 3856

Search completed: October 30, 2002, 10:50:56
Job time : 52.6131 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:10:53 : Search time 1044.68 seconds
(without alignments)
3888.820 Million cell updates/sec

Title: US-09-836-602-6
Perfect score: 1509
Sequence: 1 MPRGTALCLSLPSPGMS.....LVTSGALLVLGITGYFLMN 301

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q/cn2_l/USPTO_spool/US09836602/runat_29102002_101144_3812/app_query.fasta_1.1429
-DB=EST -OFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptlo -NORM=ext -HICAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602.qcgn.1.1.2453 -runat_29102002_101144_3812 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
No.							
1	1509	100.0	977	9	AL548050	AL548050	
2	1409	93.4	919	9	AL513812	AL513812	

3	1153	76.4	735	9	AL598821	AL598821	DKF2p313K
4	1122	74.4	761	9	AU141253	AU141253	AU141253
5	1096	72.6	752	9	AU141111	AU141111	AU141111
6	1078	70.9	749	9	AU121093	AU121093	AU121093
7	1069.5	70.9	768	9	AU141883	AU141883	AU141883
8	1040	68.9	1129	10	BI769224	BI769224	6030600051
9	994	65.9	804	10	EG675028	EG675028	602621258
10	991.5	65.7	754	9	AU135182	AU135182	602621258
11	907.5	60.1	1001	10	EG256232	EG256232	602369196
12	792	52.5	717	9	AU121843	AU121843	602369196
13	755.5	50.1	733	9	AU141125	AU141125	602369196
14	714	47.3	825	9	AU139201	AU139201	602369196
15	685.5	45.4	903	10	BI824944	BI824944	603033307
16	627	41.6	464	10	BM146840	BM146840	603033307
17	627	41.6	551	10	T80274	T80274	603033307
18	624	41.4	477	10	BM146845	BM146845	603033307
19	610	40.4	839	10	BF178239	BF178239	601807770
20	609	40.4	780	9	AI806138	AI806138	601807770
21	571	37.8	603	9	AA673625	AA673625	601807770
22	553	36.6	657	9	AL547643	AL547643	601807770
23	541	35.9	382	10	C18678	C18678	601807770
24	514	34.1	464	9	AA792789	AA792789	601807770
25	496	32.9	473	10	BF830285	BF830285	601807770
26	490	32.5	486	9	AI893233	AI893233	601807770
27	468	31.0	424	10	W65699	W65699	601807770
28	467	30.9	477	9	AI371138	AI371138	601807770
29	416	27.6	805	10	EG864235	EG864235	602797477
30	411	27.2	806	10	BE913124	BE913124	601668112
31	405	26.8	997	9	AL547847	AL547847	601668112
32	390	25.8	397	10	BF993687	BF993687	601668112
33	365	24.2	223	10	H72215	H72215	601668112
34	361.5	24.0	675	9	BB665591	BB665591	601668112
35	350	23.2	558	9	AW656254	AW656254	601668112
36	338	22.4	297	10	BF471335	BF471335	601668112
37	313.5	20.8	551	9	AW910263	AW910263	601668112
38	302	20.0	263	9	AW839633	AW839633	601668112
39	297.5	19.7	1081	10	BE916866	BE916866	601668112
40	287.5	19.1	609	9	BB659396	BB659396	601668112
41	281.5	18.7	575	9	AA907722	AA907722	601668112
42	273.5	18.1	653	9	BB642054	BB642054	601668112
43	257	17.0	549	10	BE948643	BE948643	601668112
44	253	16.8	248	9	AA311473	AA311473	601668112
45	249	16.5	474	10	BE847358	BE847358	601668112

ALIGNMENTS

RESULT 1
AL548050

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

FEATURES

source

AL548050 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1034YP22 5
prime, mRNA sequence.

AL548050.1 GI:12882692
EST.

human.
Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 977)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.

Full-length cDNA libraries and normalization
Unpublished (2001)

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers
1..977

/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="CS0D1034YP22"

/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 259 a 272 c 231 g 214 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 2 64e-125 Length: 977
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AL548050 (1-977)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 40 ATGCCGGGGCTGGACCGCGCTTTGCTGCTGAGTTTGCTGCTTCTGGGTTTCATGAGT 99
Qy 21 LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 100 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAATGTT 159
Qy 41 SerThrAsnValSerTyrglnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 160 TCTCAATGATGCTCTACCAAGAACTACACACTAGTACCTTGGAGTACCAGGCTG 219
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 220 CACCCTGTGCTCAACATGCAATGAGCCACACAAACATCACAGAAACAGACAGTCAAA 279
Qy 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
Db 280 TTCACTACTACCTCTGTGATAACCTCAGTTTATGGAAACACAACTCTTCTGTCAGTCA 339
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 340 CAGACCTCTGTAATACACACAGAGTTCCACCCACCCACAGTTTCACTTCCAGAGACA 399
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 400 ACCTTGAAGCTAGCCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCCTAGC 459
Qy 141 LeuAlaThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIleLys 160
Db 460 CTTGCAACATCTCCACATAACCCCTATACATCATCTTCTCTATCTTAAGTGACATCAAG 519
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 520 GCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATGACTAGGGCATCTGCTGGAG 579
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 580 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAAGACAGGGGAGGCGCTGGCCCGA 639
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 640 GTGCTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGCGCCAGGATGCTCCCTGCTC 699
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 700 CTTGCCCCAGTCTGAGGTGAGGCCCTCAGTGTCTACTGCTGTGGTGGCCACACAGAA 759
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260

Db 760 ATTTCCAGCAACTCCAACTATGAAAAGCACCACAACTCTGACCTGAAAAGCTGGGATC 819
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrsGlnLysThrLeuLeu 280
Db 820 CTAGATTTCACTGAGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAGACCTGATT 879
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrrPheLeuMet 300
Db 880 GCACCTGTCACCTCGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCTCATG 939
Qy 301 Asn 301
Db 940 AAT 942

RESULT 2
AL513812 919 bp mRNA linear EST 13-FEB-2001
LOCUS AL513812 LTI_NFL006_PL2 Homo sapiens cDNA clone CLOBA0092F10 5
DEFINITION prime, mRNA sequence.

ACCESSION AL513812 GI:12777306
VERSION AL513812.1
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 919)

AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

Location/Qualifiers

1..919

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CLOBA0092F10"

/tissue_type="placenta"

/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 246 a 253 c 213 g 198 t 9 others

ORIGIN

Alignment Scores:

Pred. No.: 2 25e-116 Length: 919

Score: 1409.00 Matches: 286

Percent Similarity: 97.61% Conservative: 0

Best Local Similarity: 97.61% Mismatches: 7

Query Match: 93.37% Indels: 1

DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AL513812 (1-919)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20

Db 40 ATGCCGGGGCTGGACCGCGCTTTGCTGCTGAGTTTGCTGCTTCTGGGTTTCATGAGT 99

Qy 21 LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

Db 100 CTTGACAAACACGGTACTGCKACCCAGAGTTTACCACCCAGGGAACATTTTCAATGTT 159


```
Db 628 GAGTGTCTGTGGGAGGAGCAGCTGATGCTGATGCTGGGGCCCGCCAGGTATGCTNCCTGC 687
Qy 220 euLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeu 235
|||||
Db 688 TCCTTGGCCAGCTGAGGNGAGGCTCAGTGTCTACTGCTGCTTG 734

RESULT 4
LOCUS AU141253 761 bp mRNA linear EST 25-OCT-2000
DEFINITION AU141253 THYR01 Homo sapiens cDNA clone THYR01000286 5', mRNA
sequence.
ACCESSION AU141253
VERSION AU141253.1 GI:11002774
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 761)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000286"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 197 a 214 c 180 g 167 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 9.51e-91 Length: 761
Score: 1122.00 Matches: 222
Percent Similarity: 98.68% Conservative: 2
Best Local Similarity: 97.80% Mismatches: 3
Query Match: 74.35% Indels: 0
DB: Gaps: 0

US-09-836-602-6 (1-301) x AU141253 (1-761)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 73 ATGCGCGGGGCTGACCGCGCTTGTGCTGAGTTGCTGCTTCTGGTTTCATGAGT 132
|||||
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 133 CTTGACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGAACATTTCAAATGTT 192
|||||
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 193 TCTACAATGTATCTTACCAAGAACTACAAACACCTAGTACCTTGGAAATACAGGCTG 252
|||||
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
|||||
Db 253 CACCCTGTGCTCAACATGCAATGAGGCCACACACAGCATCACAGAAACACAGTCAA 312
|||||
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
```

```
Db 313 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 372
|||||
Qy 101 GlnThrSerValIleSerThrValPheThrThrProIaAsnValSerThrProGluThr 120
|||||
Db 373 CAGACCTCTGTAATCAGCAGAGTGTTCACCAACCCACCAACGTTTCAACTCCAGAGACA 432
|||||
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 433 ACCTTGAAGCTAGCCCTGCTCACCCTGGAATGTTTCAGACCTTTCACCACTAGCAGTAC 492
|||||
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 493 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTCTTCTTCTTCTTCT 552
|||||
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 553 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAAGTGAATTTGACTCAGGGCATCTCGCTGGAG 612
|||||
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 513 CAAAATAAGACCTTCCAGCTGTGGGAGTTTAAAGAGACAGGGGAGGGGCGCTGGCCCGA 672
|||||
Qy 201 ValLeuCysGlyGluGlnGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 673 GTGCTGTGTGGGAGGAGCANGCTGATGCTGCTGGGCCCAAGTATGCTCCCTGNTC 732
|||||
Qy 221 LeuAlaGlnSerGluValArg 227
|||||
Db 733 CTTGCCCACTCTGANGTGAAG 753
|||||

RESULT 5
LOCUS AU141111 752 bp mRNA linear EST 25-OCT-2000
DEFINITION AU141111 THYR01 Homo sapiens cDNA clone THYR01000038 5', mRNA
sequence.
ACCESSION AU141111
VERSION AU141111.1 GI:11002632
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1..752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000038"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 197 a 208 c 176 g 168 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 2.01e-88 Length: 752
```

Score: 1096.00 Matches: 218
Percent Similarity: 98.20% Conservativeness: 0
Best Local Similarity: 98.20% Mismatches: 4
Query Match: 72.63% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AU141111 (1-752)

Qy	1	MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	65	ATGCGCGGGGTGACCGCGTGTGCTGCGAGTTGCTGCCCTTCGGGTTCATGAGT	124
Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	125	CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTT	184
Qy	41	SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu	60
Db	185	TCTACAAATGTTATCTTACCAAGAACTACAAACCTAGTACCTTTGGAAGTACCGACCTG	244
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
Db	245	CACCTGTGTCTCAACATGGCAATGAGGCCACACAACATCACAGAAACGACAGTCAAA	304
Qy	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
Db	305	TTTCATCTTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA	364
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	365	CAGACCTCTGTAATCAGCACAGTGTACACCCCGGCAACGTTTCAACTCCAGAGACA	424
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	425	ACCTTGAAGCCTAGGCTGTACCTGGAATGTTTTCAGACCTTTCAACCACTAGCAGTAC	484
Qy	141	LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
Db	485	CTTGCAACATCTCCACATAACCCCTATACATCATCTTCTCCCTATCCTAAGTGACATCA	544
Qy	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	545	GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCCTG	604
Qy	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	605	CAAAATAAGACCTTCAGCTGTGCGGAGTTTAAAGAGCAGGGGAGGGGCTGGCCCGA	664
Qy	201	ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	665	GTGCTGTGTGGGAGGAGCAGCTGATGCTGATGCTGGGGCCCAAGTATTGCTCTGCTC	724
Qy	221	LeuAla 222	
Db	725	CTTGCC 730	

RESULT 6
LOCUS AU121093 749 bp mRNA linear EST 19-OCT-2000
DEFINITION AU121093 HEMBB1 Homo sapiens cDNA clone HEMBB1002065 5', mRNA
sequence.
ACCESSION AU121093
VERSION AU121093.1 GI:10936328
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 749)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project

JOURNAL
COMMENT

Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5', & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES
source

Location/Qualifiers
1..749
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBB1002065"
/clone_lib="HEMBB1"
/tissue_type="whole embryo, mainly body"
/dev_stage="embryo, 10 weeks"
/note="Vector: pME18SFL3"
BASE COUNT 195 a 211 c 173 g 166 t 4 others
ORIGIN

Alignment Scores:

Pred. No.: 8.22e-87 Length: 749
Score: 1078.00 Matches: 214
Percent Similarity: 96.83% Conservativeness: 0
Best Local Similarity: 96.83% Mismatches: 7
Query Match: 71.44% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AU121093 (1-749)

Qy	1	MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	73	ATGCGCGGGGTGACCGCGTGTGCTGCTGAGTTTGCCTTCTGGGTTCATGAGT	132
Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	133	CTTGACAACAACGGTACTGCTACCCAGAGTTACCTTACCCAGGGAACATTTTCAAATGTT	192
Qy	41	SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu	60
Db	193	TCTACAAATGTTATCTTACCAAGAACTACAAACCTAGTACCTTGAAGTACCGCTG	252
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
Db	253	CACCTGTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA	312
Qy	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
Db	313	TTTCATCTTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA	372
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	373	CAGACCTCTGTAATCAGCACAGTGTTCACCCCGGCAACGTTTCAACTCCAGAGACA	432
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	433	ACCTTGAAGCCTAGGCTGTACCTGGAATGTTTTCAGACCTTTCAACCACTAGCAGTAC	492
Qy	141	LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
Db	493	CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCCCTATCTTCTTATGATGACATCA	552
Qy	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	553	GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCANGGCATCTGCCTG	612
Qy	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	613	CAAAATAAGACCTTCAGCTGTGCGGAGTTTAAAGAGCAGGGGAGGGGCTGGCCCGA	672

```

QY 201 ValLeuCysGlyGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 673 GTGCTGTGGGGAGAACANGCTGATGCTGATGCTGGGGCCCAAGTATGCTCTGTCTCC 732
QY 221 Leu 221
|||
Db 733 TTG 735

RESULT 7
AUI141883 768 bp mRNA linear EST 25-SEP-2000
LOCUS AUI141883 THYR01 Homo sapiens cDNA clone THYR01001399 5', mRNA
DEFINITION
ACCESSION AUI141883
VERSION AUI141883.1 GI:11003404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamanoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01001399"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/notes="Vector: pME18SF13"
BASE COUNT 198 a 215 c 181 g 171 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 4.93e-86 Length: 768
Score: 1069.50 Matches: 217
Percent Similarity: 95.61% Conservative: 1
Best Local Similarity: 95.18% Mismatches: 10
Query Match: 70.87% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AUI141883 (1-768)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 72 ATGCCGGGGGCTGGACCGGCTTTGCTGCTGAGTTTGCTGCTTCTGGGTTTCATGAGT 131
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 132 CTTGACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGAGAACATTTCAAAATGTT 191
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 192 TCTACAATGTATCTACCAAGAACTACAACTAGTACCTTGGAGTACAGGCTG 251
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
|||||

```

```

Db 252 CACCTGTGTCTCAACATGGCAATGAGCCACAAACACACAGAAACACAGAGTCAAA 311
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 312 TTCACATCTACTCTGTGATAACCTCAGTTTATGGAACACAAAACCTCTTCTGTCAGTCA 371
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 372 CAGACCTCTGTAATCAGCACAGTGTTCACACCCAGCCAGCGTTTCAACTCCAGAGACA 431
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 432 ACCTTGAAGCCTAGCTGCTCACCCTGGAATGTTTCAGACCTTTCAACCACTAGCACTAGC 491
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 492 CTTGCAACATCTCCCACTTAAACCTATACATCTTCTCTATCTTCCCTAAGTGACATCAAG 551
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 552 GCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATGACTCAGGGCATCTGCCTGGAG 611
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 612 CAAATAAGACCTTCAGCTGTGCGAGTTTAAAGAAGACAGAGGGAGGCGCTGCCCGCA 671
QY 201 ValLeuCysGlyGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 672 GTGCTGTGGGGAGGAGAACANGCTGATGCTGATGCTGGC--CCAGTATGCTTCTCTGCTC 729
QY 221 LeuAlaGlnSerGluValArgPro 228
|||||
Db 730 CTTGCCAATCTGANGTGAAGGCCT 753

RESULT 8
BI769224 1129 bp mRNA linear EST 25-SEP-2001
LOCUS 603060051f1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5209333 5',
DEFINITION mRNA sequence.
ACCESSION BI769224
VERSION BI769224.1 GI:15760802
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1129)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM11525 row: e column: 14
High quality sequence stop: 765.
FEATURES
source
1..1129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5209333"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/notes="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed)

```

upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."

BASE COUNT	292 a	317 c	279 g	241 t
ORIGIN				

Alignment Scores:		
Pred. No.:	3,79e-83	1129
Score:	1040.00	247
Percent Similarity:	81.70%	Conservative: 12
Best Local Similarity:	77.92%	Mismatches: 31
Query Match:	68.92%	Indels: 27
DB:	10	Gaps: 7

US-09-836-602-6 (1-301) x BI769224 (1-1129)

QY	1	MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
DB	55	ATCCCGGGGGCTGGACCGCGCTTTCCTGCTGAGTTGCTGCTCTGGGTTCATGAGT	114
QY	21	LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
DB	115	CTTGACACACACGGTACTGCTACCCACAGAGTTACCTPACCCAGGGAACATTTTCAAAATGTT	174
QY	41	SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu	60
DB	175	TCACAAATGATATCCTACCAAGAAACTACAAACCTTAGTACCTTGGAAAGTACCAGCCGT	234
QY	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
DB	235	CACCCCTGTGTCTCAACATGGCAATGAGGCCACAAACAACATCACAGAAACGACAGTCAAA	294
QY	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
DB	295	TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAAACACAAACTCTTCTGTCCAGTCA	354
QY	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
DB	355	CAGACCTCTGTAATCAGCACAGTGTTCACACCCACGCCAACGTTTCAACTCCAGAGACA	414
QY	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
DB	415	ACCTTGAGGCTAGCTGTCTACCTGGAAATGTTTACAGACTTTCAACCACTAGCATAGC	474
QY	141	LeuAlaThrSerProThr-LysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
DB	475	CTTGCAACATCTCCCACTAAAGCCCTATACATCATCTTCTCCTATCTTAAGTGACATCAA	534
QY	160	sAlaGluIleLysCysSerGlyIle-ArgGluVal-LysLeuThrGlnGlyIleCysLeu	179
DB	535	GGCAGAAATCAAAATGTTTCAGGCATCACAGAGAAGTGACAATGACTCAGGGCATCTGCCTG	594
QY	180	GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla	199
DB	595	GACCAAAATAGACCTCCAGCTGTGCGGAGTTTACAGAGCACAGGGAGAGGGCCCTGGCC	654
QY	200	-ArgValLeuCys--GlyGluGluGlnAlaAspAlaGlyAla-Gln---ValC	217
DB	655	CCGAGTGTCTGTGCGGCGAGGAGCAGCGCTGATGCTGCTGGGGCCCCAGCGTATGC	714
QY	217	ysSerLeuLeuLeuAlaGlnSerGlu-ValArgPro-GlnCysLeuLeuLeuValLeuAl	236
DB	715	TCCCGGTGGTCTTGCCAGCTTGAGCGGTGAGCCCTCCAGTGTCTACTGCTGGGTCTTGC	774
QY	236	a-----AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGln	252
DB	775	GCCCCTAGGAACGGGAATTTGCCCCAGCAACAGCTCCCAAGCTTTATGAANAAGCACCCA	834
QY	252	nSerAspLeuLysLysLeu-----GlyIleLeuAspPheThrGluGlnAspValAlaSe	270
DB	835	ATCTGAGCCCTGAACCAAGCTGGGGGAATCCATAGAATTTTCGCTGACAGAGATGTTCCGAC	894

Qy	270	rHisGlnSerTyrSerGlnLysThrLeuIleAla	-----Le 282
	:	:	
Db	895	AAGCATAGC-----AAGAAACTATTTCCTCCCAAGAGAGACCCCTGATTGGCACCT	945
Qy	282	uValThr---SerGlyAlaLeuLeuAlaValLeuGly	293
Db	946	GGAGACCCCTCGGGAAGCCCTGCTGGATGTACTTGGG	982
RESULT	9		
LOCUS	BG675028		
DEFINITION	60262125BF1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4746591 5' ,	804 bp	linear EST 01-MAY-2001
ACCESSION	BG675028	nrna sequence.	
VERSION	BG675028.1	GI:13906424	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 804)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D.		

Email: crabbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLML at:

```

http://image.llnl.gov
Plate: LRAM10595 row: d column: 16
High quality sequence stop: 686.
Location/Qualifiers
1. .804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cbone="IMAGE:4746591"
/clone_lib="NCI_CGAP_Skn3"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: This is a NCI_CGAP Library."
219 a 217 c 191 nt 177 t
BASE COUNT

```

BASE COUNT	219 a	217 g	191 g	177 t
ORIGIN				
Alignment Scores:				
Pred. No.:	3.11e-79	Length:		804
Score:	994.00	Matches:		229
Percent Similarity:	91.41%	Conservative:		5
Best Local Similarity:	89.45%	Mismatches:		10
Query Match:	85.87%	Indels:		12
DB:	10	Gaps:		2

US-09-836-602-6 (1-301) x BG675028 (1-804)

Qy	1	MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	46	ATGCCGGGGCTGGACCGCCTTGGCTGCTGAGTTTGGCTTCCTGGGTTTCATGAGT	105
Qy	21	LeuAspAsnAnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	106	CTTGACAAACAACGGTACTGCTACGCCAGAGTACCTACCCAGGGAACATTTCAAATGTG	165
Qy	41	SerThr-AsnValSerTyrGlnGlu--ThrThrProSerThrLeuGlySerThrSe	59
Db	166	TCTCAAAGTGTATCCTACCAAGAAAGTACACGACCTAGTACCTTGTAAAGTACCAG	225
Qy	59	rLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVa	79

```

Db 226 CCTGCACCCCTGTCTCAACATGGCAATGAGGCCACAAACAAACATCACAGAAACGACAGT 285
Qy 79 llySerPheThr-SerThrSerValIleThrSerValThrGlyAsnThrAsnSerSerValG 99
Db 286 CAANTTCAGCATCTACTCTGTGATACCTCAGTTTATGGAACAAACACTCTCTGTCC 345
Qy 99 lnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProG 119
Db 346 AGTCACAGACCTCTGTAAATCAGCACAGTGTTCACCAACCCAGCAACGTTTCAACTCCAG 405
Qy 119 luThrThr-LeuLysProSerLeuSerProGlyAsnValSerAsp-LeuSerThrThr-S 138
Db 406 AGACAAGCCCTTGAAGCCTAGCCTGTCCACCTGGAATGTTTCAGACCCCTTTCAACCACTAG 465
Qy 138 erThrSerLeuAlaThrSerProThrLysProThrLysProThrLysSerSerProIleLeuSerA 158
Db 466 GCCTAGCCTTGCACATCTCCACCTAAACCTATACATCTCTCTATCTCTAAGT 525
Qy 158 spleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleC 178
Db 526 ACATCAAGCAGAAATCAATGTTTCAGGCATCAGAGAGTGAAATTCAGCTCAGGGCATCT 585
Qy 178 ysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyL 198
Db 586 GCCTGGAGCAAAATAAGACTCCTCAGCTGTGCGGAGTTTAAGAAGGACAGGGGAGGGAC 645
Qy 198 euAlaArgValLysCysGlyGluGluGln-AlaAspAlaAsp-AlaGlyAlaGlnVal-- 216
Db 646 TGCCCGAGTGTGTGTGGGAGGAGCAGAGCTGTGCTGATAGCTGGGGGCCAGGATATA 705
Qy 217 -CysSerLeuLeuAlaGlnSer-GluValArgProGlnCysLeuLeu-LeuValLeu 235
Db 706 GCTCCCTGCTACCTTGCCCGCAGCTAGAGGTGAGGCTCAGTGTCTACTGACTGGTCCCTG 765
Qy 236 -AlaAsnArgThrGluIleSerSerLys 244
Db 766 GGCAACAGACAGACATTTTCAGCAAAA 793

RESULT 10
AUI35182 754 bp mRNA linear EST 24-OCT-2000
LOCUS AUI35182 PLACE1 Homo sapiens cDNA clone PLACE1001404 5', mRNA
DEFINITION AUI35182
ACCESSION AUI35182
VERSION AUI35182.1 GI:10995721
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 754)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
TITLE Unpublished (2000)
JOURNAL Genomics Laboratory
COMMENT Contact: Takao Isogai
1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers
1..754
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1001404"
source
```

```

/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: pME18SPL3"
BASE COUNT 201 a 210 c 172 g 166 t 5 others
ORIGIN
Alignment Scores:
Pred. No.: 4,76e-79 Length: 754
Score: 991.50 Matches: 208
Percent Similarity: 90.60% Conservative: 4
Best Local Similarity: 88.89% Mismatches: 12
Query Match: 65.71% Indels: 10
DB: 9 Gaps: 3
US-09-836-602-6 (1-301) x AUI35182 (1-754)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 48 ATCCCGCGGGGCTGGACCGCGCTTGTGCTGAGTTGTGCTGCTTCTGGGTTTCATGAGT 107
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 108 CTTGACAAACACCGTACTGCTACCCAGAGTTACTACCCAGGGGAACATTTTCAATGTT 167
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 168 TCTACAAATGTATCCTTACCAGAAACTACAAACACCTAGTACCCTTGAAGTACCAGCCTG 227
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 228 CACCCCTGTCTCAACATGGCAATGAGCCACAAACATCACAGAAACGACAGTCAA 287
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 288 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 347
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 348 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCAACTTTCACTCCAGAGACA 407
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 408 ACCTTGAAGCCTACCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCCTAGC 467
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
Db 468 CTTGCAACATCTCCCACTAAACCTTATACATCATCTTCTCTATCTTAAAGTGAATCAAG 527
Qy 161 Ala-GluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuG1 180
Db 528 GCAAGAAATCAAAATGTTTCAGGCATCAGAGAAAGTGAAATTCAGCTCAGGGCATCTGC 587
Qy 180 uGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu---GlyLeuAl 199
Db 588 GCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAGAAGGACAGAGGGAAGGCGCTGGCC 647
Qy 199 aArgVal---LeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSe 218
Db 648 CCAGTTGTGTGTGGGGAGGAGCAAGGCTGATGCC-----TG 686
Qy 218 rLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeu 231
Db 687 ATTGCTGGGGGCCCAAGTTANNNGCTCCCTGCTGCTCCCTT 726
RESULT 11
BG256232 1001 bp mRNA linear EST 13-FEB-2001
LOCUS BG256232
DEFINITION BG256232
ACCESSION BG256232
VERSION BG256232.1 GI:12766048
KEYWORDS EST.
SOURCE human.
```

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1001) NIH-MGC http://mgc.ncl.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: DCTD/DTF cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAMi0305 row: p column: 23 High quality sequence start: 10 High quality sequence stop: 579. Location/Qualifiers 1..1001 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4477318" /clone_lib="NIH_MGC_91" /tissue_type="adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /notes="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library." 283 a 287 c 257 g 174 t
FEATURES	source
BASE COUNT	

```

BASE COUNT      283 a      287 c      257 g      174 t
ORIGIN
Alignment Scores:
Pred. No.:      2.44e-71      Length:      1001
Score:          907.50      Matches:      209
Percent Similarity: 89.87%      Conservative: 4
Best Local Similarity: 88.19%      Mismatches: 15
Query Match:    80.14%      Indels:      10
DB:             10      Gaps:        1
US-09-836-602-6 (1-301) x BG256232 (1-1001)

```

QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 274 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTT 333
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 334 TCTACAAATGTATCCTACCAAGAAACTACAACACCTAGTACCCCTTGAAGTAGTACGACCTG 393
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
|||||
Db 394 CACCCTGTGCTCAACATGGCAATGAGGCCACCAACAAACATCAGAAAGACGAGTCAAA 453
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 454 TTCACATCTACCTCTCTGATACCTCAGTTTATGGAACACACAACTCTTCTGTGCCAGTCA 513
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 514 CAGACCTCTGTATATCAGCACAGTGTTCACACCCCAAGCAACGTTTCAACTCCAGAGACA 573
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 574 ACCTTGAAGCCTAGCCTGTACCTTGAAATGTTTCAGACCTTTCAACCACTAGCACTAGC 633
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 634 CTTGCAACATCTCCACTAAACCCATATCA-TCATCTTCNTCTATNCTAAGTGACATCAAG 692
QY 161 AlaGluIleLysCysSerGlyIle 168
|||||
Db 693 GCAGAAATCAAAATGTTTCANGCATC 716
RESULT 13
AUI141125 733 bp mRNA linear EST 25-OCT-2000
LOCUS AUI141125 THYR01 Homo sapiens cDNA clone THYR01000063 5', mRNA
DEFINITION sequence.
ACCESSION AUI141125
VERSION AUI141125.1 GI:11002646
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 733)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. 733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000063"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 199 a 212 c 168 g 150 t 4 others
ORIGIN

Alignment Scores: 6.73e-58 Length: 733
Pred. No.: 755.50 Matches: 157
Score: 94.61% Conservative: 1
Percent Similarity: 94.01% Mismatches: 8
Best Local Similarity: 50.07% Indels: 2
Query Match: 9 Gaps: 1
DB: 1
US-09-836-602-6 (1-301) x AUI141125 (1-733)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 233 ATGCCCGGGGCTGGACCGCTTTGCTTGTGAGTTTGCCTTCTGGGTTTCATGAGT 292
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 293 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTTACCCAGGGAACATTTTCAAATGTT 352
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 353 TCTACAAATGTATCCTTACCAAGAAACTACAACACCTAGTACCCCTTGAAGTAGTACCGCTG 412
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
|||||
Db 413 CACCCTGTGCTCAACATGGCAATGAGGCCACCAACAAACATCAGAAAGACGAGTCAAA 472
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 473 TTCACATCTACCTCTCTGATACCTCAGTTTATGGAACACACAACTCTTCTGTGCCAGTCA 532
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 533 CAGACCTCTGTATATCAGCACAGTGTTCACACCCCAAGCAACGTTTCAACTCCAGAGACA 592
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 593 ACCTTGAAGCCTANCTGTACCTTGAAATGTTTCAGACCTTTCAACCACTAGCACTACCT 652
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 653 TCGACATCTCC---ACTAAACCCATATCA-TCATCTTCNTCTATCTCTAAGTGACATCAAG 708
QY 161 AlaGluIleLysCysSerGly 167
|||||
Db 709 GCAGAAATCAAAATGTTTCAGGC 729
RESULT 14
AUI139201 825 bp mRNA linear EST 25-OCT-2000
LOCUS AUI139201 PLACE1 Homo sapiens cDNA clone PLACE1010139 5', mRNA
DEFINITION sequence.
ACCESSION AUI139201
VERSION AUI139201.1 GI:11000722
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 825)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of

Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

source
1. .825
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1010139"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: pME18SFL3"
BASE COUNT 227 a 243 c 185 g 165 t 5 others
ORIGIN

Alignment Scores:
Pred. No.: 4,21e-54 Length: 825
Score: 714.00 Matches: 148
Percent Similarity: 95.54% Conservatives: 2
Best Local Similarity: 94.27% Mismatches: 4
Query Match: 47.32% Indels: 3
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AU139201 (1-825)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 301 ATCCGCGGGCTGGACCGGCTTGTCTGCTGAGTTTGTCTGCTTCTGGTTCATGAGT 360
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 361 CTTTGACAAACACGGTACTGTCTACCCAGAGTTTACCTACCCAGGGAACATTTTCAAAATGTT 420
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 421 TCTACAATATGATCTTACCAAGAACTACAAACCTAGTACCTTGGAACTACCAAGCTG 480
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 481 CACCCTGTGTCTCAACATGGCAATGAGGCCACCAACAAACATCAGAAACGACAGTCAAA 540
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 541 TTCACATCTACCTCTGTGATTAACCTCAGTTATATGGAACACAAACTCTTCTGCCAGTCA 600
QY 101 Gln-ThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 601 CAGACCTCTGTAATCAGCAGAGTGTTCANCAACCCAGCCCAACGTTTCAACTCCAGAGAC 660
QY 120 rThrLeuLysProSerLeuSerProGly-AsnValSerAspLeuSerThrThrThrS 140
Db 661 AACCTTGAAGCCTTAACCTGTCACTGCTGAGAAATGTTTCAAACTTTTAACCACTANCACTA 720
QY 140 erLeuAlaThrSerProThr-LysProTyrThrSerSerPro 154
Db 721 GCCTTGCAACATCTCCCACTAAAACCCATATACATCATCTTCTCCT 765

RESULT 15
BI824944
LOCUS BI824944 903 bp mRNA linear EST 04-OCT-2001
DEFINITION 603033307F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174593 5',
mRNA sequence.

ACCESSION BI824944

VERSION BI824944.1 GI:15936494

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 903)

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11434 row: n column: 02
High quality sequence start: 6
High quality sequence stop: 587.

FEATURES

source

1. .903
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5174593"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."

BASE COUNT 291 a 262 c 203 g 147 t

ORIGIN

Alignment Scores:
Pred. No.: 1.72e-51 Length: 903
Score: 685.50 Matches: 165
Percent Similarity: 81.43% Conservatives: 6
Best Local Similarity: 78.57% Mismatches: 26
Query Match: 45.43% Indels: 13
DB: 10 Gaps: 2
US-09-836-602-6 (1-301) x BI824944 (1-903).

QY 1 MetProArgGlyTrpThrAlaLeuCys-LeuLeu-SerLeuLeuProSerGlyPheMet- 19
Db 99 ATCCGCGGGCTGGACCGGCTTGTCTGCTGAGATTTGTCTGCTTCTGGTTCATGA 158
QY 20 SerLeuAspAsnAsnGly-ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAs 39
Db 159 GATCTTGACAAACACGGATACTGTACCCAGAGTTACCTACCCAGGGAACATTTTCAAA 218
QY 39 nValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSe 59
Db 219 TGTTCCTACAAATCTATCTTACCAGAAACTACAAACCTAGTACCTTGGAACTACCA 278
QY 59 rLeuHisProValSerGlnHisGlyAsnGluAlaThr-ThrAsnIleThrGluThrThr 79
Db 279 CCTGCACCTGTGTCTCAACATGGCAATGAGGCCACCAACAAACATCAGAAACGACAG 338
QY 79 allysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValG 99
Db 339 TCAAAATTCACATCTACCTCTGTGATACCTCAGTTATGGAACACAAACTCTTCTGTGCC 398
QY 99 InSerGlnThrSerValIleSerThrValPheThrThrPro-AlaAsnValSerThrPro 118
Db 399 AGTCACAGACCTCTGTAATCAGCAGACAGTGTTCACCACCCAGACCAACGTTTCAACTCCA 458
QY 119 GluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThr--- 137
Db 459 GAGACAAACCTTGAAGCCCTAGCCTGTCACTGGAATGTGTGAGACCTTTGCAAAACACTAG 518
QY 138 SerThrSerLeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSer 157
Db 519 CACATACCATGACAAACATCTCCCACTAAACCCCTATACATCATCTTCTCTATCTTAAGT 578
QY 158 AspIleLysAla--GluIleLysCysSerGly-IleArg-----GluValLysLeuThr 174

```

|||||  : : : : : |||||  : : : : : |||||  : : : : : |||||  : : : : : |||||
Db  579 GACATCAAGGCCAGAAACTCAAAATGTCCGGCAATCAGACAAAGCGAAACTGACTCCGCG 638
QY  175 GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGlu-PheLysLysAspAr 194
Db  639 CCATCGGCCNAGCGCAGCAAACTTAAGACCTCCAGACTGGCGGAGGCACACGAAAGGACAA 698
QY  194 gGlyGluGly 197
Db  699 GGGCACAGGG 708
: |||
: |||
: |||
```

Search completed: October 30, 2002, 10:48:00
Job time : 1048.68 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:08:08 ; Search time 1543.01 Seconds
(without alignments)
4082.221 Million cell updates/sec

Title: US-09-836-602-6
Perfect score: 1509
Sequence: 1 MPRGWTALCLSLLLPSGFMS.....LVTSGALLAVLGITGVFLMN 301

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+lp2n.model -DEV=xlp
-Q=/cgn2_1/USPFO_Spool/US09836602/runat_29102002_101144_3785/app_query.fasta_1.1429
-DB=GenEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602 -CGN_1_1_3800 -runat_29102002_101144_3785 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MMAB -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgtg_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1509	100.0	2463	9 S53910	S53910 CD34-glycop
2	1509	100.0	2615	9 HUMCD34HS	M81104 Human CD34
3	1509	100.0	2657	6 AX333753	AX333753 Sequence
4	1509	100.0	2657	9 S53911	S53911 CD34-glycop
5	899	59.6	2956	4 CFU49457	U49457 Canis famil
6	850	56.3	2690	4 AF461503	AF461503 Sus scrof
7	839	55.6	1065	6 AR078454	AR078454 Sequence
8	805.5	53.4	978	10 S69301	S69301 CD34-cell s
9	805.5	53.4	1149	10 S69299	S69299 CD34-cell s
10	805.5	53.4	1260	10 S69293	S69293 CD34-stem c
11	805.5	53.4	2398	10 BC006607	BC006607 Mus muscu
12	728	48.2	581	9 AF202879	AF202879 Homo sapi
13	707.5	46.9	2453	4 AB021662	AB021662 Bos tauru
14	681.5	45.2	2420	4 AF259378	AF259378 Capra hir
15	617	40.9	160771	9 HS8812	AF259378 Human DNA
16	617	40.9	182742	2 AC084393	AC084393 Homo sapi
17	617	40.9	185162	2 AL365178	AL365178 Homo sapi
18	617	40.9	212535	2 AL356275	AL356275 Homo sapi
19	409	27.1	254	9 HUMCD34S3	M81940 Human CD34
20	313.5	20.8	98347	2 AL513203	AL513203 Mus muscu
21	313.5	20.8	166333	2 AL513470	AL513470 Mus muscu
22	311	20.6	183	9 HUMCD34S2	M81939 Human CD34
23	260	17.2	157	9 HUMCD34S5	M81942 Human CD34
24	233	15.4	235	6 AX247364	AX247364 Sequence
25	214	14.2	165	9 HUMCD34S7	M81944 Human CD34
26	172.5	11.4	3124	5 MMTHRAD1	Y13978 Gallus gall
27	167	11.1	80272	5 AL591180	AL591180 Zebrafish
28	166	11.0	89232	5 AL591175	AL591175 Zebrafish
29	162.5	10.8	55069	8 YSCH9196	U11583 Saccharomyc
30	162.5	10.8	184427	14 EHVD20824	U20824 Equine herp
31	160.5	10.6	253273	2 AL669823	AL669823 Mus muscu
32	158.5	10.5	183358	9 AC019041	AC019041 Homo sapi
33	156.5	10.4	2161	5 XELFIMC1X	L02115 Frog integu
34	155.5	10.3	167390	9 AC007263	AC007263 Homo sapi
35	155.5	10.3	182534	10 AC074046	AC074046 Mus muscu
36	154	10.2	3053	8 SCYNR0444W	Z11659 S.cerevisia
37	154	10.2	3137	8 YSCAAGLCS	M60590 S.cerevisia
38	153.5	10.2	167254	9 CNS05TDS	AL357093 Human chr
39	153.5	10.2	172789	2 AC104020	AC104020 Homo sapi
40	153	10.1	165227	2 AC098014	AC098014 Rattus no
41	153	10.1	185337	2 AC095762	AC095762 Rattus no
42	153	10.1	286060	2 AL589862	AL589862 Homo sapi
43	152.5	10.1	196553	2 AC073946	AC073946 Mus muscu
44	152	10.1	12412	3 CELH43E16	AC006669 Caenorhab
45	152	10.1	131274	2 AC096869	AC096869 Rattus no

ALIGNMENTS

RESULT 1

S53910
LOCUS S53910
DEFINITION CD34-glycoprotein expressed in lymphohematopoietic progenitor cells
(alternatively spliced) [human, UT7 cells, mRNA, 2463 nt].
ACCESSION S53910
VERSION S53910.1 GI:264766
KEYWORDS human UT7 cells.
SOURCE

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2463)
AUTHORS Nakamura,Y., Komano,H. and Nakauchi,H.
TITLE Two alternative forms of cDNA encoding CD34
JOURNAL Exp. Hematol. 21 (2), 236-242 (1993)
MEDLINE 93146100
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI glibsq 124007] from the original journal article.
This sequence comes from Fig. 1A.

FEATURES
source
1..2463
/location="Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene
1..2463
/gene="CD34"
CDS
91..1248
/gene="CD34"
/note="glycoprotein expressed in lymphohematopoietic
progenitor cells; This sequence comes from Fig. 1A"
/codon_start=1
/product="CD34"
/protein_id="AAB25222.1"
/db_xref="GI:264767"
/translation="MLVRRGARGPRMGWTALCLLSLLPSGFMSLDNNGTATPELP
TQTFNSVNTSVSYQFTTPSTLGSLSHPVSOHQNEATNITETTVKFTSTSVITSV
YGNATSSVQSQTSVITPANNVSTPETTLKPSLSPGVNSDLSTTSLSATSPKP
YSSPFLSDQKLLAQCSEVREKVLQGLQKNTSSCAEFKKGEGELARVLGCEE
QADADAGVSKLLAQSEVRPQCLLLANRTEISSKLQMKKHQSDLKGLIDFT
EDVASHQSYQKTLIALTVSGALLAVLGITGYFLMNRSSPTGERLGEDPYITENG
GGQGYSSGPGTSPAQGKASVNRGVAQENGTGQATSRNGHSARQHVAVDEL"
BASE COUNT 589 a 709 c 580 g 585 t
ORIGIN

Alignment Scores:
Pred. No.: 9,67e-92 Length: 2463
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-836-602-6 (1-301) x S53910 (1-2463)

QY 1 MetProArgGlyThrPheAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCCGGGGCTGGAGCCGCGCTTGTCTGCTGAGTTGCTGCTCTCTGGGTTTCATGAGT 186
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACACACACGGTACTGCTACCCAGAGTTACCTACCCAGGAGACATTTTCAATGTT 246
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTACAAATGTATCCTACCAAGAAACTACACACCTAGTACCTTGGAGTACACGCTG 306
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 307 CACCCCTGTGCTCAACATGCAATGAGGCCACACAAACATCACAGAAACGACAGTCAA 366
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATACTCAGTTTATGGAACACAACTCTCTCTGCTCAAGTCA 426
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTAATCAGCAGAGTGTTCACCACCCAGCAACGTTTCAACTCCAGAGACA 486
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 487 ACCTTAAGCCTAGCTGTACCTCGAGAAATGTTTTCAGACCTTCAACCACTAGCAGTAGC 546

QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 547 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTAAAGTACATCAAG 606
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 607 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAAGTGAATTTGACTCAGGGCATCTCCCTGGAG 666
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 667 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGACAGAGGGAGGCGCTTGGCCGA 726
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 727 GTCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGCCCATGATGCTCCCTGCTC 786
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 787 CTTGCCCAGTCTCAGGTGAGGCGCTCAGTGTCTACTGCTGCTTGGCCACAGACAGAA 846
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 847 ATTTCAGCAAACTCCAACCTTATGAAAAGCACCAATCTGACCTGAAAAGCTGGGATC 906
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 907 CTAGATTTCACCTCAGCAAGATGTTGCAAGCCACACAGAGCTATTCCCAAGACCCCTGATT 966
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 967 GCACCTGTCACCTCGGAGCCCTGCTGGCTGTCTTGGGCATCCTGCTATTTCTCTGATG 1026
QY 301 Asn 301
Db 1027 AAT 1029
RESULT 2
LOCUS HUMCD34HS 2615 bp mRNA linear PRI 02-NOV-1993
DEFINITION Human CD34 mRNA, complete cds.
ACCESSION M81104.1
VERSION M81104.1 GI:180108
KEYWORDS CD34; hematopoietic stem cell surface antigen; sialomucin.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2615)
AUTHORS Simmons,D.L., Satterthwaite,A.B., Tenen,D.G. and Seed,B.
TITLE Molecular cloning of a cDNA encoding CD34, a sialomucin of human
hematopoietic stem cells
JOURNAL J. Immunol. 148, 267-271 (1992)
MEDLINE 92091783
FEATURES
source
1..2615
/location="Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
294..1415
/codon_start=1
/product="CD34"
/protein_id="AAA03181.1"
/db_xref="GI:180109"
/translation="MPRGWTALCLLSLPSGFMSLDNNGTATPELPTGTFNSVNTNV
SYQETTPSTLGSLSHPVSOHQNEATNITETTVKFTSTSVITSVYGNNTSSVQSQT
ABIKCSGIREVKLTQGLCEQNTSSCAEFKKGEGELARVLGCEQADADAGQVCS
LLLAQSEVRPQCLLLANRTEISSKLQMKKHQSDLKGLIDFTEDDVASHQSYSQ
KTLLALVTSGALLAVLGITGYFLMNRSSPTGERLGEDPYITENGGGQGYSSGPGTS
PEAQGKASVNRGAAQKNGTQATSRNGHSARQHVAVDEL"
BASE COUNT 616 a 763 c 627 g 609 t
ORIGIN
Alignment Scores:

[illegible]

Qy	81	PheThrSerThrSerValIleThrSerValTyGlyAsnThrAsnSerSerValGlnSer	100
Db	536	GTCAATCTACCTCTGAGATCAACCCCTAAACGCGCTGGAGCATGAATCTCTGTCTGTCAGTCG	595
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	596	CAGACCTCTTTAGCTATCAGGTATCTTTTACCCCAACCAACTTTTCAACTCTCAAGTGTG	655
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	656	ACCTTGGAGCCGACCTGTCTACCTGGAATGTTTCGGATCCCCCTACACAGCAGCACCCAGC	715
Qy	141	LeuAlaThrSerProThrLysProThrLysProThrSerSerSerPro	154
Db	716	CTTGTGACATCCCCCGAATATTATACATCACTTTCTTCCTACCCCAAGTAGAAATGAC	775
Qy	155	IleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThr	174
Db	776	ACCCCAAGTACCATCAGGAGAAATCAATGTTCCGGAGTCAAGAAGTGAATTTGAAC	835
Qy	175	GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArg	194
Db	836	CAAGGTATCTGCCTAGAGCTAAATGAGACCTCCAGCTGTGAGGACTTTTAAGAAAGATAAC	895
Qy	195	GlyGluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAla	214
Db	896	GAAGAAAACTGACCCCAAGTCCTGTGT- ----GAGAAGGAGCAGCTGAGGCTGGGCCC	949
Qy	215	GlnValCysSerLeuLeuLeuAlaGlnSerGluValAlaArgProGlnCysLeuLeuVal	234
Db	950	GGGTGTGTCCTCCCTCTCTGCGCCAGCTCTGAGGTGAGGCCCTCACTGCCTGCTGCTGTC	1009
Qy	235	LeuAlaAsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAsp	254
Db	1010	TGCGCCAAACAAACAGAACCTTTTCACTAACTCACTCTCTGAGAAAGCACCAGCTGCAC	1069
Qy	255	LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyr	274
Db	1070	CTGAAAAAGCTGGGGATCCGAGACTTCACTCAACAAGATGTTGGGAGCCACAGAGCTAT	1129
Qy	275	SerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIle	294
Db	1130	TCCGCAAGACCCCTGATGACCTGGTCACCTCAGGGATCCTCTGGCTGTCTTGGGCACC	1189
Qy	295	ThrGlyTyrPheLeuMetAsn	301
Db	1190	ACTGGTTACTCTCGATGAAC	1210
RESULT 6			
AF461503			
LOCUS	AF461503	2690 bp	mRNA linear
DEFINITION	Sus scrofa CD34 antigen mRNA, complete cds.		
ACCESSION	AF461503		
VERSION	AF461503.1	GI:18308137	
KEYWORDS			
SOURCE			
ORGANISM	Sus scrofa		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		
	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
REFERENCE			
AUTHORS	1 (bases 1 to 2690)		
TITLE	Sun, J. and Butler, J.E.		
JOURNAL	Cloning, sequencing and expression of swine CD34 gene		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 2690)		
TITLE	Sun, J., Wang, J. and Butler, J.E.		
JOURNAL	Direct Submission		
FEATURES	Submitted (20-DEC-2001) Microbiology, University of Iowa, 51 New		
	Road, Iowa City, IA 52242, USA		
	Location/Qualifiers		
	1..2690		
	/organism="Sus scrofa"		
	/db_xref="taxon:9823"		
	/tissue_type="mesenteric lymph nodes"		

```

CDS      298...1467
         /note="cluster of differentiation antigen 34"
         /codon_start=1
         /product="CD34 antigen"
         /protein_id="AAU67838.1"
         /db_xref="GI:18308138"
         /translation="MPRGWNTTCLLSLLPSGFTAVNSSTVASTLPTAAGSTPTGPATA
GTATSAISDISESSVPQTSIAVTSSATINFTSEVTLQPFPFGNVSDPLYNSTSPA
TSRTSLVPETVNSVPQTHKGVCAQIKVKLTQICLERNETSCCEKKFKDNKEKLMO
RSPSTVSPPTPGSHKGVBKCAQIKVKLTQICLERNETSCCEKKFKDNKEKLMO
ILCGQQAEPAGPCVCILLAQASVKPHCLLVLANGTELSKFLLELKHQSELREMSI
CHFSQKDVRSHQSYSKRKTIALYVTSIGILLAVLGITGYLLMNRASVPTGBRLGEDPYI
TENGGQGVSYSGPAPGAQSVTRGAQENGTGQATSRNGHSARQPMVAADTEL"
BASE COUNT   547 a    808 c    738 g    597 t
ORIGIN
Alignment Scores:
Pred. No.:          9.28e-48           Length:                2690
Score:              850.00             Matches:            187
Percent Similarity: 68.63%             Conservative:        34
Best Local Similarity: 58.07%           Mismatches:         75
Query Match:       56.33%               Indels:             26
DB:                4                   Gaps:               6

US-09-836-602-6 (1-301) x AF461503 (1-2690)

Qy      1 MetProArgGlyThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db      298 ATGCCGGGGCTGGACCACGCCTCGCTGCTGAGTTTGTCGCCCTCGGTTTCACAGCT 357
Qy      21 LeuAspAsnAnglyThrAlaThrProGluLeuPro----- 32
Db      358 GTG---AACAGCTCGACTGTGTCTCCACCTTGCCAACATGCCGCTGGGTCAACTCCCACC 414
Qy      33 -----ThrGlnGlyThr-----pheSerAsn 39
Db      415 GGGCCAGCTACTCGAGGAGCACAGCTATCACCGGGTCAGCTATCTCAGACATATCTTCACCT 474
Qy      40 ValSerThrAsnValSerTyrgIngluThrThrThrProSerThrLeuGlySerThrSer 59
Db      475 GTTTCTACAATATATCCACGAGGAA--ACCACACCATGCTTTTCGAAAGTCCGACG 531
Qy      60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVal 79
Db      532 CTCACACACTGCTCTCAGGCGCAGCAGTGGGACCACCGCTAGCCATCTCAGGCCCTACAGTT 591
Qy      80 LysPheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGln 99
Db      592 AATTTCATGTCACCTCGGGGTCCACCTGCTCCCCGAAAACCGTGAATCTCTCTGTCGAC 651
Qy      100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
Db      652 CCTCAGACCTCT---CTAGCCACAGTGTCTCCGCCACCATCACTTTACAACTTCAGAG 708
Qy      120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 139
Db      709 GTGACCTCGAGCCGACGACGCTTCCCAGGAAATGTTTTCAGACCCCTCTCAACACAGTACC 768
Qy      140 SerLeuAlaThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIle 159
Db      769 AGCCCTCGGAGATCCCCACGACGCCCTACACATCATCTCTCTCTCTCTCTCTCTCTCTCT 828
Qy      160 LysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Db      829 AAGGGGAAGTCAANTGTGCCCAANTCAAAGAGGTGAATTTGACCAAGGTATCTCGCCTG 888
Qy      180 GluGlnAsnLysThrSerSerCysAlaLeuPheLysLysAspArgGlyGluGlyLeuAla 199
Db      889 GAGCGAAATGAGACCTCCGGCTCGGAGAGCTTTAAGAAGGACAAATGGAGAGAGATTGATG 948
Qy      200 ArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
Db      949 CAATCTCTGTGTGGCAGGACGAGCTGTGGCGGGCCAGGG-----GTGTGCTCTCTTG 1002

```



```
Db 361 AATTTCATGTCTACCTCGCGGCTCACCTCGTCCCGGAAACCGTTAACTCTCTCTGTCCAG 420
Qy 100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
Db 421 CCTCAGACCTCT---CTAGCCACAGCGTCTCCGCCACCATCACTTTACAACTTCAGAG 477
Qy 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrSerThr 139
Db 478 GTGACCTCGCAGCCAGCAGCTTCCAGGAATCTTTCAGACCCCTCTACACAGTACC 537
Qy 140 SerLeuAlaThrSerProThrLysProThrLysProThrLysProThrLysProThr 159
Db 538 AGCCCTCGAGATCCCGACACGAGCCCTTACACATCATCTCTCTACCCCGAGTAGCCAC 597
Qy 160 LysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyLeuCysLeu 179
Db 598 AAGGGGAGTCAATGTGCCAATCAAGAGGTGAATGACCAAGGTATCTGCCTG 657
Qy 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Db 658 GAGCGAATGAGACCTCGCGCTCGAGAGTTTAAGAGGACAATGGAGAGAGTGTATG 717
Qy 200 ArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
Db 718 CAAATCTGTGTGGCGAGCAGCTGAGCGCGGCGCAGGG-----GTGTGCTCCTTG 771
Qy 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThr 239
Db 772 CTCCTTGCCCAATCTGAGGTGAACCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831
Qy 240 GluIleSerSerLysLeuGlnMetLysLysHisGlnSerAspLeuLysLysLeuGly 259
Db 832 GAATTAGCACCAAGTCTCTCTGCGAAGAGCACCAGCTGACTGAGAGAGATGAGC 891
Qy 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerLysThrLeu 279
Db 892 ATCCAAAATCTCGAAAAGATGTTAGGAGCCACAGCAGCTACTCCCGAAGACCTTG 951
Qy 280 IleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyThrPheLeu 299
Db 952 ATTGCATGGTCACCTCGGGGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1011
Qy 300 MetAsn 301
Db 1012 ATGAAC 1017
RESULT 8
S69301
LOCUS
DEFINITION CD34=cell surface antigen (alternatively spliced, clone 2) [mice,
stromal cell line PA-6, mRNA Partial, 978 nt].
ACCESSION S69301
VERSION S69301.1 GI:495715
KEYWORDS
SOURCE Mus sp. stromal cell line PA-6.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Suda,J., Sudo,T., Ito,M., Ohno,N., Yamaguchi,Y. and Suda,T.
TITLE Two types of murine CD34 mRNA generated by alternative splicing
JOURNAL Blood 79 (9), 2288-2295 (1992)
MEDLINE 9233983
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gidsq 145459] from the original journal article.
This sequence comes from Fig. 3.
FEATURES
source 1..978
/db_xref="taxon:10095"
gene 1..978
/partial
/gene="CD34"
```

```
CDS
1..978
/gene="CD34"
/note="cell surface antigen; This sequence comes from Fig.
4"
/codon_start=1
/product="CD34"
/protein_id="AAB22109.1"
/db_xref="GI:495716"
/translaton="MOVHRDTRAGLLLPWVVALCLMSLLHNLNLTSAATTSTOGTIS
PSVPTNESVEENITSSIPGSTSHYLIQDSKSTTPALSETWVNTVTSGLTPSSGTPH
TFSPQPTSGIILPTSDSLSTSEMTKWSLPSINVSIDSPNNSSEFMTSPTPIYAT
SSAPSAIKGEIKSGIREVRLAQICLSEASCEEFKEKGEDILCLCKEAE
ADAGASVCSLLLAQSEVRPCLLMANSTELPSKLQMEKHQSDLRKLGIOFQNKQD
IGSHQSVSRKTLIALVTSGVLLAILGTGTGFLMNRSSWSPTGRLELEP"
BASE COUNT 261 a 249 c 228 g 240 t
ORIGIN
Alignment Scores:
Pred. No.: 2,36e-45 Length: 978
Score: 805.50 Matches: 177
Percent Similarity: 71.15% Conservative: 40
Best Local Similarity: 58.03% Mismatches: 77
Query Match: 53.38% Indels: 11
DB: 10 Gaps: 4
US-09-836-602-6 (1-301) x S69301 (1-978)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 37 CTGCCATGGCGCTGGTAGCTCTCTGCTGCTGATGCTCTGCTG-----CAT 81
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 82 CTAATAACTTGTACTTCTGCTACCAGGAGACTTCTACACAGGAATAATCCCATCAGTT 141
Qy 41 SerThrAsnValSerValThrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 142 CCTACCAATGAGTCTGTGAGGAAATATACATCATGATCCCTGGAGGTACCAGCAC 201
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 202 TACTTGATCTATCAGCAGCAGCAGTAGTAAAGCACCAGCAGCATCTCAGAGCATATGTTCAAC 261
Qy 81 PheThrSerThrSerValIleThrSerValThrGlyAsnThrAsnSerSerValGlnSer 100
Db 262 TTTACAGTTACCTCTGGGATCCCTTCCAGGCTCTGGAACCTCCACACACTTTTTCACAACA 321
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 322 CAGACTTCCCAACTGGCATACTGCTTACTACTTCCAGACAGTATTTCCACTTCAGAGATG 381
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 382 ACCTGGAAGTCCAGCCTGCCATCTATAAATGTTTCTGATTATTCGCTTAATAATAGCAGC 441
Qy 141 Leu--AlaThrSerProThrLysProThr-----ThrSerSerSerProIleLeu 156
Db 442 TTTGAGATGATCACCACCCAGCAGCATATGCTTACACATCATCTCTGCTGCCG----- 495
Qy 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db 496 AGTCCATTAAGGAGGAATCAATGCTCTGGAATCCGAGAAGTGAGGTTGCCCGCAGGT 555
Qy 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 556 ATCTGCTGGAACAAAGTGAAGCATCTAGTTCTGAGGAGTTTAAAGAGGAAGAGAGAA 615
Qy 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGluVal 216
Db 616 GATCTAATTAATTAATCTGTGTGAAGGAGGAGGCTGAGGCTGAGTGGTCTAGTGTCTC 675
Qy 217 CysSerLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
```

```

Db 676 TGCTCCTGCTTCTAGCCCAAGTCTGAGGTTAGCCCTGAGTGTTGCTGATGGCTTGGGC 735
QY 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
||| |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 736 AATAGCACAGAACTTCCAGCAACTCCAGCTTATGGAAAGACCAATCTGACTTGAGA 795
QY 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 796 AAGCTGGGATCCAAAGCTTCAATAAACAAGATATCGGAGCCACCAAGAGCTATTCCCGA 855
QY 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 856 AAGACTCTTATGCTATGGTTCCTGAGGTTCTGCTGGCCATCTTGGGCAACCATGGT 915
QY 297 TyrPheLeuMetAsn 301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 916 TATTTCCTGATGAAC 930

RESULT 9
S69299
LOCUS
DEFINITION S69299 1149 bp mRNA linear ROD 03-JUN-1994
[mice, stromal cell line PA-6, mRNA partial, 1149 nt].
ACCESSION S69299
VERSION S69299.1 GI:495713
KEYWORDS Mus sp. stromal cell line PA-6.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1149)
AUTHORS Suda,J., Sudo,T., Ito,M., Ohno,N., Yamaguchi,Y. and Suda,T.
TITLE Two types of murine CD34 mRNA generated by alternative splicing
JOURNAL Blood 79 (9), 2288-2295 (1992)
MEDLINE 92239883
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 145456] from the original journal article.
This sequence comes from Fig. 3.
FEATURES
source
1. .1149
/organism="Mus sp."
/db_xref="taxon:10095"
1. .1149
/partial
/gene="CD34"
1. .1149
/gene="CD34"
/note="cell surface antigen; This sequence comes from Fig.
3"
/codon_start=1
/product="CD34"
/protein_id="AAB22108.1"
/db_xref="GI:495714"
/translation="MQVHRDTRAGLLLPWRVVALCLMSLLHLLNLTSAATSTSGIS
PVPNSESVEENITSSIGSTSHYLIYQDSKTPAISETMYNTVFTVIGPSGSGTPH
TFSPQSTPTGLPTSTISSTSEMTKSLPSINVSDYSPNNFSFEMTSPTEPYAT
SSSAIKGELKCGIRVLAOGICLSESPASCEPKKQKEDLIQLCEKEAE
ADAGASVCSLLLAQSEVRPECLIMVLANSTELPSKLQMEKHQSDRLKLGIOENKQD
IGSHOSYSRKTLIALVTSGVLLAIGTGYFLNRRNSPTGERLGEDPYPTENGGO
GYSSGPGASPEPQGANVTRGAQENGTOATSRNGHSARQHVVADTEL"
BASE COUNT 307 a 293 c 285 g 264 t
ORIGIN

Alignment Scores:
Pred. No.: 2,9e-45 Length: 1149
Score: 805.50 Matches: 177
Percent Similarity: 71.15% Conservative: 40
Best Local Similarity: 58.03% Mismatches: 77
Query Match: 53.38% Indels: 11
DB: 10 Gaps: 4

US-09-836-602-6 (1-301) x S69299 (1-1149)

```

```

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
::: ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37 CTGCCATGGCGCTGGCTAGCTCTCTGCCATGATGAGTCTGCTG-----CAT 81
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 82 CTAATAACTTGACTTCTGCTACCAAGGAGACTTCTACACAAGGAATATCCCATCAGTT 141
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 142 CCTACCAATGAGTCTCTGAGGAAATATACATCTAGCATCCTCGAAGTACCAAGCCAC 201
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
::: ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 202 TACTTGATCTATCAGGACAGCAGTAAGACCAACCAAGCCATCTCAGAGACTATGGTCAAC 261
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 81 PheThrSerThrValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 TTTACAGTTACCTCTGGGATCCCTTCAGGCTCTGGAACTCCACACACTTTTTCACAACCA 321
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 CAGACTTCCCAACTGCCTACTGCTACTTCTACAGAGTATTTCCACTTCAGAGATG 381
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 382 ACCTGGAAGTCCAGCCTGCCATCTATAAATGTTTCTGATTATTCGCTATAATAGCAGC 441
QY 141 Leu--AlaThrSerProThrLysProTyr-----ThrSerSerSerProIleLeu 156
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 442 TTTGAGATGATCATCACCCAGGACCATATGTTACACATCATCTTCTGCTCG----- 495
QY 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 496 AGTGCCATTAAGGAGAAATCAATGCTCTGGAATCCGAGAAGTGAAGTTGGCCAGGT 555
QY 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 556 ATCTGCTGGAACATAAGTGAAGCATCTAGTTGTGAGAGAGTTTAAAGAGAAAAGGAGAA 615
QY 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 616 GATCTAATTCAAATACTGTGTGAAAGGAGGAGGCTGAGGCTGATGCTGCTGCTAGTGC 675
QY 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 676 TGCTCCTGCTTCTAGCCAGTCTGAGTTAGGCTGAGTGTGTTGCTGATGGCTTTGGCC 735
QY 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 736 AATAGCACAGAACTTCCAGCAACTCCAGCTTATGGAAAGACCAATCTGACTTGAGA 795
QY 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 796 AAGCTGGGATCCAAAGCTTCAATAAACAAGATATCGGAGCCACCAAGAGCTATTCCCGA 855
QY 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 856 AAGACTCTTATGCTATGGTTCACCTCTGAGGTTCTGCTGGCCATCTTGGCACCACCATGGT 915
QY 297 TyrPheLeuMetAsn 301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 916 TATTTCCTGATGAAC 930

RESULT 10
S69293
LOCUS
DEFINITION S69293 1260 bp mRNA linear ROD 03-JUN-1994
[CD34-stem cell antigen (mice, bone marrow, mRNA, 1260 nt)].
ACCESSION S69293
VERSION S69293.1 GI:495708
KEYWORDS Mus sp. bone marrow.
SOURCE Mus sp.
ORGANISM Mus sp.

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1260)

Brown, J., Graves, M. F. and Molgaard, H. V.
The gene encoding the stem cell antigen, CD34, is conserved in mouse and expressed in haemopoietic progenitor cell lines, brain, and embryonic fibroblasts

Int. Immunol. 3 (2), 175-184 (1991)

91223042
GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 145442] from the original journal article.

This sequence comes from Fig. 2.

Location/Qualifiers

1..1260

/organism="Mus sp."
/db_xref="taxon:10095"

1..1260
/gene="CD34"

24..1172
/gene="CD34"

/note="stem cell antigen; This sequence comes from Fig. 2"

/codon_start=1
/product="CD34"

/protein_id="AAB19246.1"

/db_xref="GI:495709"

/translation="MOVHRTDTRAGLLPWRVALCLMSLLHLLNLTTSATTETSTQGIS
PSVPTNESVBEITSSIPGSTSHVLYQDSSTKTPALSETWVNFVTSGIPSGSGTH
TFQSPQSPGILPTTSDSISTSEMTWKSLSIPNSDSYFNSSFWFTPTPTAT
SSASAIKEIGKCSGIREVRLAAGICLSEASEEKEFKEDGDIQLCERKEAE
ADAGASVCSLLAQSEVREPLCLWLANSTELPSKLQMEKHQSDRLKLGTSFNKQD
IGSHQSRSLTGLVGLLAILGTGYFLMNRWSPTGERLGDPTVTENGSGQ
GYSSGSGPETQCKANVTGRGAQENGTQATSRNGHSRQHVVADTEL"

BASE COUNT 325 a 321 c 320 g 294 t

ORIGIN

Alignment Scores:

Pred. No.: 3.26e-45 Length: 1260
Score: 805.50 Matches: 177
Percent Similarity: 71.15% Conservativeness: 40
Best Local Similarity: 58.03% Mismatches: 77
Query Match: 53.38% Indels: 11
DB: 10 Gaps: 4

US-09-836-602-6 (1-301) x S69293 (1-1260)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20

Db 60 CTGCCATGGCGCTGGTGGTACCTCTGCCTGATGAGTCTGCTG-----CAT 104

Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

Db 105 CTAATAACTTGCCTCTGCTACACGGAGACTTCTACACAGGAATATCCCCATCAGTT 164

Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60

Db 165 CCTACCAATGAGTCTGTTGAGGAAATATCATATCATGACCTCCCTGGAAGTACCGCCAC 224

Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80

Db 225 TACTTGATCTATCAGACAGCAGTGAAGACACACACGACCATCTCAGAGACTATGTCAC 284

Qy 81 PheThrSerThrValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100

Db 285 TTTACAGTTACTCTGGGATCCCTTCAGGCTCTGGAATCCACACATTTTTCACACCA 344

Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120

Db 345 CAGACTTCCCAACGCGCATCTGCTACTTCTCAGACAGTATTTCCACITTCAGAGATG 404

Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140

Db 405 ACCTGGAAGTCCAGCTGCCATCTATAATGTTCTGTATTTATTCGCTTATAATAATAGCAGC 464

Qy 141 Leu---AlaThrSerProThrLysProTyr-----ThrSerSerSerProfileLeu 156
Db 465 TTTGAGATGATCATCCACCCAGGCATATGCTTACACATCATCTTCTGCTCCG----- 518
Qy 157 SerAspileLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db 519 AGTGCCATTAGGAGGAATAATGCTCTGGAATCCGAGAAGTGAGTTGGCCAGGGT 578
Qy 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 579 ATCTGCTGGAACCTAAAGTGAAGCATCTAGTTGTGAGGAGTTTAAAGAGGAAAGGAGAA 638
Qy 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyValGlnVal 216
Db 639 CATCTAATTCAAATACCTGTGTAAAGAGGAGGCTGAGGCTGATGCTGGTGTAGTGTG 698
Qy 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db 699 TGCTCCCTGCTTCTAGCCACGCTGAGGTTAGGCTGAGTGTGCTGCTGATGCTGTGGCC 758
Qy 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
Db 759 AATAGCACAGAACTTCCAGCAAACTCCAGCTTATGGAAGACCAATCTGACTGTGAGA 818
Qy 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
Db 819 AAGCTGGGATCCAAAGCTTCAATAAACAAGATATCGGAGGCCACGAGCTATTTCCGA 878
Qy 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
Db 879 AAGACTCTTATTGCTTGGTACCTCTGAGTCTGCTGGCCTGCTGCTGGCACCACCTGCT 938
Qy 297 TyrPheLeuMetAsn 301
Db 939 TATTTCTGTGATGAAC 953

RESULT 11
BC006607 2398 bp mRNA linear ROD 12-JUL-2001
LOCUS Mus musculus, Similar to CD34 antigen, clone MGC:11720
DEFINITION IMAGE:3966337, mRNA, complete cds.
ACCESSION BC006607
VERSION BC006607.1 GI:13879273
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2398)

Strausberg, R.
Direct Submission

Submitted (27-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk

Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 17 Row: n Column: 8

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES		Location/Qualifiers	
source		1..2398	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/clone="MGC:11720 IMAGE:3966337"	
		/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."	
		/clone_lib="NCL_CGAP_Mam1"	
		/lab_host="DH10B"	
		/note="Vector: pCMV-SPORT6"	
CDS		62..1210	
		/codon_start=1	
		/product="Similar to CD34 antigen"	
		/protein_id="AAH06607.1"	
		/db_xref="GI:13879274"	
		/translation="MQVHRDTRAGLLPWRVALCLMSLLHNLNLTSAFTTSTOGIS PSVPTNESVEENITSSIPGTSHTYLIYQDSKTTTPAISRTMVFNTVTSIPSGSGTPH TFSPQTSPTGILPTTSDISISTSEMTWKSLLPSINVSDYSPNNSFEMTSPTEPVAYT SSASATKGEIKKCSIREVRLAOGICILELSPASSCEEFKEKGGDLIQLCEKEAE ADAGASVSLLLAQSREVRPECLLMVLANSTELPSKLQIMEKHQSDLRKLGIGSFNKQD IGSHQSYSKRTIALIVTSGVLLAILGTGTYFLMNRKSWSPTEGERGEDPYITENGGGQ GYSSGPGASPEPQGKANVTRGAENGTCQATSRNGHSARQHVYVADTEL"	
BASE COUNT		623 a 599 c 551 g 625 t	
ORIGIN			
Alignment Scores:			
Pred. No.:	7.42e-45	Length:	2398
Score:	805.50	Matches:	177
Percent Similarity:	71.15%	Conservative:	40
Best Local Similarity:	58.03%	Mismatches:	77
Query Match:	53.38%	Indels:	11
DB:	10	Gaps:	4
US-09-836-602-6 (1-301) x BC006607 (1-2398)			
QY	1	MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer	20
Db	98	CTGCCATGGCGCTGGTAGTCTCTGCCTGATGAGTCTGCTG-----CAT	142
QY	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	143	CTAAATAACTGTACTCTGTACACCGAGACTTCTACACAAGGAATATCCCATCAGTT	202
QY	41	SerThrAsnValSerThrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu	60
Db	203	CCTACCAATGAGTCTGTGAGGAAAAATATCACATCTAGCATCCCTGGAGTACCGCCAC	262
QY	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
Db	263	TACTTGATCATCGGACAGCAGTGAACACACACGCCATCTCAGAGACTATGGTCAAC	322
QY	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
Db	323	TTTACAGTTACCTCTGGATCCCTCTCAGGCTCTGGAACCTCCACACACTTTTTCACAACCA	382
QY	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	383	CAGACTTCCCAACTGGCCTACTGCTACTTTCAGACAGATATTTCACAGATG	442
QY	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	443	ACCTGGAAGTCCAGCTGCCATCTATAAATGTTTCTGATATTTCGGCTAATAATAGCAGC	502
QY	141	Leu---AlaThrSerProThrLysProTyr-----ThrSerSerSerProIleLeu	156
Db	503	TTTGAGATGACATCACCCACCGAGCCATATGCTTACACATCATCTTCTGCTCG-----	556
QY	157	SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly	176
Db	557	AGTGCCATTAAAGGAGAAATCAAAATGCTCTGGAAATCCGAGAGTGAAGTGGTGGCCAGGGT	616

QY	177	IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu	196
Db	617	ATCTGCTGGAACTAAGTGAAGCATCTAGTTGTGAGGAGTTTAAAGAGGAAAGGGAGAA	676
QY	197	GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal	216
Db	677	GATCTAATTCAAATCTGTGTGAAAAGGAGGAGCTGAGGCTGATGCTGTGTAGTGTCT	736
QY	217	CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla	236
Db	737	TGCTCCTCTGCTTCTAGCCAGTCTGAGGTGAGCCTGAGTGTGCTGATGCTGTGGCC	796
QY	237	AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys	256
Db	797	AATGACGACAGAACTTCCCAGCAAACTCCAGCTTATGGAAGACCACTATCGACTTGAGA	856
QY	257	LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerThrSerGln	276
Db	857	AGCTGGGATCCAAAGCTTCAATAACAAGATATCGGAGCCACCAGAGCTATTCCTCGA	916
QY	277	LysThrIleuIleAlaLeuValThrSerGlyAlaLeuAlaValLeuGlyIleThrGly	296
Db	917	AGACTCTTATTGCTATGTCACCTCTGGAGTCTGCTGCCATCTTTGGCCACCACTGGT	976
QY	297	TyrPheLeuMetAsn	301
Db	977	TATTCTCTGATGAAC	991
RESULT	12		
AF202879		581 bp	mRNA
LOCUS			
DEFINITION			Homo sapiens hematopoietic progenitor cell antigen CD34 precursor
ACCESSION			AF202879
VERSION			AF202879.1
KEYWORDS			GI:6503195
SOURCE			human.
ORGANISM			Homo sapiens
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE			1 (bases 1 to 581)
JOURNAL			Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.
REFERENCE			Direct isolation and cDNA sequencing of mRNA from human
AUTHORS			hematopoietic progenitor cell antigen CD34
TITLE			Unpublished
JOURNAL			2 (bases 1 to 581)
REFERENCE			Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.
AUTHORS			Direct Submission
TITLE			Submitted (08-NOV-1999) Medicine, Hematology and Oncology, Sickie
JOURNAL			Cell Center, Medical College of Georgia, 15th Street, AC-1000,
FEATURES			Augusta, GA 30912, USA
source			Location/Qualifiers
			1..581
			/organism="Homo sapiens"
			/db_xref="taxon:9606"
			/chromosome="1"
			/map="1q32"
			<1..>581
gene			/gene="CD34"
CDS			<1..>581
			/gene="CD34"
			/codon_start=2
			/product="hematopoietic progenitor cell antigen CD34
			precursor"
			/protein_id="AA14634.1"
			/db_xref="GI:6503196"
			/translation="ILSDIKAEIKCSGIREVKLTQICLEQNKTSSCAEFKKDRGEGL
			ARVLGCEQADAGAQVCSLLLAQSEVRPQCLLVLANRTEISKLQMLKKHQSLDK
			KGLILDFEQDVAHQSYSKRTIALIVTSGALLAVLGITGYFLMNRKSWSPTEGERLGE
			DPYYTENGGGQGYSSGPGTSPCAQCKASVNRGA"
BASE COUNT			150 a 146 c 171 g 114 t
ORIGIN			

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 88L2. The true right end of clone 7H11 is at 71906 in this sequence. 8L2 is from the library RPC16 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:ppac4>.

FEATURES

Source

```

1..160771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q32.2-q32.3"
/clone="XX-D88L2"
/clone_lib="RPC16"
42..81
/note="10 copies 4 mer aaga 83% conserved"
42..75
/note="17 copies 2 mer aa 88% conserved"
89..412
/note="L1MC5 repeat: matches 7609. .7913 of consensus"
1027..11154
/note="L2 repeat: matches 2586. .2744 of consensus"
1189..1277
/note="L2 repeat: matches 2616. .2704 of consensus"
2426..2643
/note="MER58A repeat: matches 1. .224 of consensus"
misc_feature
complement(3075..3399)
/note="match: 252805 STS containing (CA) repeat"
3121..3176
/note="28 copies 2 mer gt 100% conserved; differs from
252805"
3806..3905
/note="L2 repeat: matches 2600. .2750 of consensus"
8052..8339
/note="L2 repeat: matches 1823. .2111 of consensus"
8406..8600
/note="MT1A1 repeat: matches 162. .365 of consensus"
8631..8815
/note="L2 repeat: matches 2158. .2346 of consensus"
8843..9132
/note="AluJb repeat: matches 1. .292 of consensus"
9295..9363
/note="L2 repeat: matches 2680. .2749 of consensus"
9372..9672
/note="match: GSS AQ075766 clone 2359015"
9493..9594
/note="MIR repeat: matches 96. .205 of consensus"
9619..9640
/note="11 copies 2 mer tc 100% conserved"
9671..9835
/note="MIR repeat: matches 46. .200 of consensus"
10253..10285
/note="L2 repeat: matches 2708. .2740 of consensus"
10284..10406
/note="MIR repeat: matches 64. .196 of consensus"
prim_transcript
<1347..11584
/note="match: multiple ESTs; match: AA884400 AA971811"
11846..11878

```

```

/note="MIR repeat: matches 1. .33 of consensus"
11879..12182
/note="AluY repeat: matches 1. .303 of consensus"
12183..12411
/note="MIR repeat: matches 33. .258 of consensus"
12443..12491
/note="L2 repeat: matches 2662. .2710 of consensus"
12574..12661
/note="MIR repeat: matches 178. .262 of consensus"
12756..13049
/note="AluJo repeat: matches 1. .300 of consensus"
13231..>13541
/note="match: STS G28100, G24984"
13231..>14597
/note="match: multiple ESTs; match: 3' EST T89361 clone
110082; Paired with EST T84126 matching this clone; match:
5' EST T84126 clone 110082; Paired with EST T89361
matching this clone"
13672..13875
/note="MIR repeat: matches 29. .231 of consensus"
14598..14907
/note="AluY repeat: matches 1. .310 of consensus"
14913..15214
/note="AluJo repeat: matches 7. .294 of consensus"
15376..15596
/note="L2 repeat: matches 1263. .1527 of consensus"
15606..15666
/note="MLT1F repeat: matches 478. .541 of consensus"
15667..16045
/note="L1PA3 repeat: matches 5764. .6146 of consensus"
16046..16084
/note="MLT1F repeat: matches 438. .478 of consensus"
16085..16379
/note="AluJo repeat: matches 1. .296 of consensus"
16380..16850
/note="MLT1F repeat: matches 1. .438 of consensus"
16854..17208
/note="L2 repeat: matches 1533. .1930 of consensus"
17268..17395
/note="L1MC4 repeat: matches 7854. .7977 of consensus"
17422..17485
/note="Charliel repeat: matches 79. .143 of consensus"
17509..17625
/note="L1MC4 repeat: matches 7706. .7822 of consensus"
17627..17677
/note="L2 repeat: matches 1949. .1995 of consensus"
17678..17986
/note="AluY repeat: matches 1. .311 of consensus"
17989..18227
/note="L2 repeat: matches 1995. .2257 of consensus"
18250..18508
/note="Charliel5 repeat: matches 2315. .2577 of consensus"
18510..18698
/note="Charliel5 repeat: matches 16. .196 of consensus"
18721..18935
/note="Charlielb repeat: matches 3. .219 of consensus"
18936..19243
/note="AluY repeat: matches 1. .307 of consensus"
19244..19531
/note="Charlielb repeat: matches 219. .501 of consensus"
19610..19779
/note="L1MD repeat: matches 1. .169 of consensus"
19780..20228
/note="L1MD repeat: matches 787. .1293 of consensus"
20299..20606
/note="L2 repeat: matches 2382. .2710 of consensus"
20625..20753
/note="L2 repeat: matches 2179. .2307 of consensus"
20813..20960
/note="L2 repeat: matches 2374. .2523 of consensus"
21149..21292
/note="MIR repeat: matches 110. .261 of consensus"
21334..22002

```

repeat_region

```

/misc_feature
<21938..22360
/note="L2 repeat: matches 1947..2724 of consensus"
<21938..22360
/note="match: GSS AQ193501 clone 2383C13"
prim_transcript
<22003..>22108
/note="match: multiple ESTs; match: AA909597 AA906144
AA907722"
repeat_region
22573..22657
/note="MIR repeat: matches 81..160 of consensus"
22573..22657
complement(24084..25666)
/note="match: STS G08650"
misc_feature
complement(join(24102..25486,26245..26409,26724..26776,
27028..27184,35057..35137,36529..36782,37377..37559,
48563..48899))
/gene="CD34"
/note="match: M81104; X60172; match: multiple ESTs; match:
T80274 C18678 AA553071 AA311526 AA311473 AA907722 AA906144
AA064307 A1173145 AA369014 R39493 AA148360 T81557 R72478
R69703 R69740 AA039660 AA639555 R73090 R73955 W92532
R66845 AA054965 AA434483 R69214 R82715 AA188435 AA524276
W7409 H01096 T12007 AA216274 AA022916 AA528783 AA483672
AA640408 W58493 T28504 A1128488 A1149563 A1144193 A1252665
A1017002 A1160509 A1017793 AA906021 A1128776 A1150834
AA434387 AA664247 AA039661 W72884 A1262206"
/evidence=not_experimental
complement(24102..48899)
/gene="CD34"
CDS
complement(join(25301..25486,26245..26409,26724..26776,
27028..27184,35057..35137,36529..36782,37377..37559,
48563..48899))
/gene="CD34"
/note="match: SWISS-PROT:P28906"
/codon_start=1
/evidence=not_experimental
/product="CD34 (hematopoietic stem cell surface antigen)"
/protein_id="CAB36556.1"
/db_xref="GI:4455631"

```

Alignment Scores:

```

Pred. No.: 5,87e-30 Length: 160771
Score: 617.00 Matches: 145
Percent Similarity: 42.3% Conservative: 1
Best Local Similarity: 42.0% Mismatches: 1
Query Match: 40.8% Indels: 198
DB: 9 Gaps: 1

```

US-09-836-602-6 (1-301) x HS88L2 (1-160771)

```

QY 14 LeuProSerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThr 33
::: |||||
Db 37563 ATAGCTTCTGGGTTTCATGAGTCTTGACACACACGGTACTGCTACCCAGAGTTACCTACC 37504
|||
QY 34 GlnGlyThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrProSer 53
|||||
Db 37503 CAGGGAACATTTTCAATGTTTCTCAATGTTATCTTACCAAGAACTACAAACACCTAGT 37444
|||||
QY 54 ThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsn 73
|||||
Db 37443 ACCCTTGAAGTACCAGCTGCACCTGTGTCTCAACATGGCAATGAGGCCACACAAAC 37384
|||||
QY 74 IleThr----- 75
|||||
Db 37383 ATCAGAGGTAAACACAGCATTTGTGTCAGATCCCGAGAGATGCTGGTGATGCTGGGTA 37324
|||
QY 75 ----- 75
Db 37323 AAGCATTTAGGATGTTTTCAGACCGCTCCCTCTCCACAGAGGAAATATACAAGTCCT 37264
|||
QY 75 ----- 75
Db 37263 AGTATTAACTGCTGAGTATCATGCTTAGGGTGCCTGAAGTAAAGTCTTAGATAATTCTT 37204
|||
QY 75 ----- 75

```

```

Db 37203 CCTTCAGTGACAGTTTTCTCAGGCCCATGCTTTGGGAGCTGATCCTGATCAGTAATGCCT 37144
QY 75 ----- 75
Db 37143 CCATACCCCTTCTCTCATGCTGAGTCTAGCTTAGTGATAAAATAAGGATCAGAAGACTCA 37084
QY 75 ----- 75
Db 37083 ATAGAGATCTCCGATCTCTTTAAAGGAAAGAGTGGGTACAGGTCCAAACTGGGGGTC 37024
QY 75 ----- 75
Db 37023 TTTGGCTTCTCTAAGGTAGACCCAGCATCTATTTCAGTTTTCAGGACACCTCTAATACATTC 36964
QY 75 ----- 75
Db 36963 TAGATTTCTAGCTCTGGTTCCAAACAGCTTGAATGAGTTTGGTCAGGGATGGGACACGAAG 36904
QY 75 ----- 75
Db 36903 TAACTGTTAACTCCCCAAACTCCCTTTCTGTGTTGAAGTGCATGCCCATATCATGACCT 36844
QY 75 ----- 75
Db 36843 GAGATTTTGTGTATCTATGACACATGCACATGCACACCCATGTTTGGTCTCTTCCA 36784
QY 76 GluThrThrValLysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsn 95
|||||
Db 36783 GAACGACAGTCAAAATTCACATCTACCTCTGTGATAAACCTCAGTTTATGGAAACACAAAC 36724
|||
QY 96 SerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnVal 115
|||||
Db 36723 TCTTCTGTCCAGTCACAGACCTCTGTATATCAGCACAGTGTTCACCACCCAGCCACGTT 36664
|||
QY 116 SerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSer 135
|||||
Db 36663 TCAACTCCAGAGACAAACCTTGAAGCTAGCCTGTACCTGGAAATGTTTTCAGACCTTTCA 36604
|||
QY 136 ThrThrSerThrSerLeuAlaThrSerProThrLysProTyrThrSerSerProIle 155
|||||
Db 36603 ACCACTAGCAGTACGCTTGCACACATCTCCCACTAAACCCCTATACATCATCTTCTCCTATC 36544
|||
QY 156 LeuSerAspIleLys 160
|||||
Db 36543 CTAAGTGACATCAAG 36529

```

Search completed: October 30, 2002, 09:52:46
Job time : 1587.01 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:10:53 ; Search time 1096.74 seconds
(without alignments)
3888.820 Million cell updates/sec

Title: US-09-836-602-4

Perfect score: 1591

Sequence: 1 MPRGWTALCLLSLPGFMS.....YFLMNRSSWSPGTGERLELEP 316

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPFO_Spool/US09836602/runat_29102002_101144_3812/app_query.fasta_1.1429
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -NODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602 -CGN_1_1_2453 -runat_29102002_101144_3812 -NCPU=6 -ICPU=3
-NO_XUPY -NO_WMAP -LARGEOQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1545	97.1	977	9	AL548050
2	1409	88.6	919	9	AL513812

3	1153	72.5	735	9	AL598821
4	1122	70.5	761	9	AU141253
5	1096	68.9	752	9	AU141111
6	1078	67.8	749	9	AU121093
7	1069.5	67.2	768	9	AU141883
8	1040	65.4	1129	10	BI769224
9	994	62.5	804	10	BG675028
10	991.5	62.3	754	9	AU135182
11	907.5	57.0	1001	10	BG256232
12	792	49.5	717	9	AU121843
13	755.5	47.5	733	9	AU141125
14	714	44.9	825	9	AU139201
15	685.5	43.1	903	10	BI824944
16	672	42.2	780	9	AI806138
17	647	40.7	839	10	BF178239
18	627	39.4	464	10	BM146840
19	627	39.4	551	10	T802174
20	624	39.2	477	10	BM146845
21	578	36.3	473	10	BF830285
22	576.5	36.2	603	9	AA673625
23	572	36.0	486	9	AI893233
24	553	34.8	657	9	AL547643
25	543	34.1	382	10	CI8678
26	533	33.5	464	9	AA792789
27	529	33.2	424	10	W65699
28	470	29.5	997	9	AL547847
29	467	29.4	477	9	AI371138
30	416	26.1	805	10	BG864235
31	415	26.1	397	10	BF993687
32	415	26.1	558	9	AW656254
33	411	25.8	806	10	BE913124
34	365	22.9	223	10	H72215
35	361.5	22.7	675	9	B665591
36	344.5	21.7	575	9	AA907722
37	338	21.2	297	10	BF471335
38	335	21.1	248	9	AA311473
39	322	20.2	549	10	BE948643
40	316	19.9	253	9	AA311526
41	314	19.7	529	10	BM069265
42	313.5	19.7	551	9	AW910263
43	302	19.0	263	9	AW839633
44	301	18.9	420	9	AL133723
45	297.5	18.7	1081	10	BE916866

ALIGNMENTS

RESULT 1

AL548050

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

AL548050 977 bp mRNA linear EST 16-FEB-2001
AL548050.LTI_NFL006_PL2 Homo sapiens CDNA clone CSODI034YP22 5
prime, mRNA sequence.
AL548050
AL548050.1 GI:12882692
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 977)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODI034YP22"
/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"
/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 259 a 272 c 231 g 214 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 9 48e-130 Length: 977
Score: 1545.00 Matches: 311
Percent Similarity: 99.68% Conservative: 0
Best Local Similarity: 99.68% Mismatches: 1
Query Match: 97.11% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x AL548050 (1-977)

Qy 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 40 ATGCCGGGGCTGGACCGCTTTGCTGCTGAGTTTGCCTTCTGGCTTCATGAGT 99
Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 100 CTTGACAAACAGGTACTGCTACCCAGAGTTACTACCCAGGAACATTTTCAATGTT 159
Qy 41 SerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 160 TCTCAAAATGATCTCTACCAAGAACTACACACTAGTACCTTGGAGTACCAGCTG 219
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValValys 80
Db 220 CACCCTGTGCTCAACATGGCAATGAGGCCAACAAACATCACAGAAACGACAGTCAA 279
Qy 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
Db 280 TTCACATCTACCTGTGATACCTCAGTTATGGAACACAAACTCTCTGTCTCCAGTCA 339
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 340 CAGACCTCTGTAACGACACAGTCTTACCACCCAGCCACACGTTTCACTCCAGAGACA 399
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 400 ACCTTGAAGCCTAGCTGTACCTGGAAATGTTTCAGACCTTTCAACCCACTAGCATTGC 459
Qy 141 LeuAlaThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIleLys 160
Db 460 CTTGCAACATCTCCCAATAACCCATACATCATCTTCTCTATCCTAAGTGACATCAAG 519
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 520 GCAGAAATCAATGTTTACGACATCAGAGAGTGAATGACTCAGGGCATCTGGCTGGAG 579
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 580 CAAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAGACAGGGGAGGGGCTGGCCCGA 639
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 640 GTGCTGTGTGGGAGGAGACAGGCTGATGCTGATGCTGGGGCCAGGATGCTCCCTGCTC 699
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 700 CTTGCCAGTCTGAGGTGAGGCCCTCAGTGTCTACTGCTGTGTTGGCCACAGAACAGAA 759
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260

Db 760 ATTTCCAGCAATCCCACTTATGAAAAGCACCAATCTGACCTGAAAAGCTGGGGATC 819
Qy 261 LeuaspPheThrGluGlnAspValAlaSerHisGlnSerTygSerGlnLysThrLeuLeu 280
Db 820 CTAGATTTCACTGACGACAGATGTTTGCACAGCCAGCAGAGCTATTCCTCAAGACCCCTGATT 879
Qy 281 AlaLeuValThrSerGlyAlaLeuAlaValLeuGlyIleThrGlyTygPheLeuMet 300
Db 880 GCATGTGTCACTCGGAGCCCTGCTGCTGTCTTGGCATCACCTGGCTATTTCTGTATG 939
Qy 301 AsnArgArgSerTipSerProThrGlyGluArgLeu 312
Db 940 AATGCCGCACGTGGAGCCCMCA-GGAGAAAGGCTG 974

RESULT 2
AL513812 919 bp mRNA linear EST 13-FEB-2001
LOCUS
DEFINITION AL513812 LTI_NFL006.PL2 Homo sapiens cDNA clone CLOBA0092F10 5
prime, mRNA sequence.

ACCESSION AL513812
VERSION AL513812.1 GI:12777306
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 919)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source

1. 919
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBA0092F10"
/tissue_type="placenta"
/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 246 a 253 c 213 g 198 t 9 others
ORIGIN

Alignment Scores:

Pred. No.: 1.77e-117 Length: 919
Score: 1409.00 Matches: 286
Percent Similarity: 97.61% Conservative: 0
Best Local Similarity: 97.61% Mismatches: 7
Query Match: 88.56% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x AL513812 (1-919)

Qy 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 40 ATGCCGGGGCTGGACCGCTTTGCTGCTGAGTTTGCCTTCTGGCTTCATGAGT 99
Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 100 CTTGACAAACAGGTACTGCKACCCAGAGTTACTACCCAGGAACATTTTCAATGTT 159


```
Db 628 GAGTGTGTGTGGGAGGAGCGCTGATGCTGATGCTGGGCCCGCCAGGTATGCTNCCTGC 687
QY 220 euLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeu 235
|||||
Db 688 TCCTTGCCAGCTGAGGNGAGGCCTCAGTGTCTACTGTGGTCTTG 734
RESULT 4
LOCUS AU141253 761 bp mRNA linear EST 25-OCT-2000
DEFINITION AU141253 THYR01 Homo sapiens cDNA clone THYR01000286 5', mRNA
sequence.
ACCESSION AU141253
VERSION AU141253.1 GI:11002774
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 761)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000286"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 197 a 214 c 180 g 167 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 1,28e-91 Length: 761
Score: 1122.00 Matches: 222
Percent Similarity: 98.68% Conservative: 2
Best Local Similarity: 97.80% Mismatches: 3
Query Match: 70.52% Indels: 0
DB: 9 Gaps: 0
US-09-836-602-4 (1-316) x AU141253 (1-761)
QY 1 MetProArgGlyThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 73 ATCCGGGGGGCTGGACCGCGCTTGTGCTGAGTTGCTGCCCTCTGGGTTCATGAGT 132
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 133 CTTGACAAACAGGTACTGTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 192
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 193 TCTACAAATGTATCTACCAAGAACTACAAACACCTAGTACCCTTGGAGTACCAGCCTG 252
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValValys 80
|||||
Db 253 CACCCTGTCTCAACATGGCAATGAGGCCAACACATCACAGAAACGACAGTCAAA 312
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
```

```
Db 313 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 372
QY 101 GlnThrSerValIleSerThrValPheThrThrProIlaasnValSerThrProGluThr 120
|||||
Db 373 CAGACCTCTGTATTCAGCAGAGTGTTCACCACCCACCAACGTTTCAACTCCAGACACA 432
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 433 ACCTTGAAGCTAGCCCTGTCACCTGGAATGTTTCAGACCTTTCACCACTAGCATTAGC 492
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 493 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTTCTCTATCTTCAAGTGACATCAAG 552
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 553 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATGACTCAGGGCATCTGCCTGGAG 612
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 613 CAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAAGCAGCAGGAGGGGCTGGCCCGA 672
QY 201 ValLeuCysGlyGluGlnAlaAspAlaGlyAlaGlnValCysSerSerLeuLeu 220
|||||
Db 673 GTGCTGTGTGGGAGGAGCANGCTGATGCTGCTGCTGGGCCCAAGTATGCTCCTGNTC 732
QY 221 LeuAlaGlnSerGluValArg 227
Db 733 CTTGCCAGTCTGANGTGAAG 753
RESULT 5
LOCUS AU141111 752 bp mRNA linear EST 25-OCT-2000
DEFINITION AU141111 THYR01 Homo sapiens cDNA clone THYR01000038 5', mRNA
sequence.
ACCESSION AU141111
VERSION AU141111.1 GI:11002632
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1..752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000038"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 197 a 208 c 176 g 168 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 2.84e-89 Length: 752
```

Score: 1096.00 218
 Percent Similarity: 98.20%
 Best Local Similarity: 98.20%
 Query Match: 68.89%
 DB: 9

US-09-836-602-4 (1-316) x AU141111 (1-752)
 QY 1 MetProArgGlyThrAlaLeuSerLeuSerLeuProSerGlyPheMetSer 20
 Db 65 ATGCCGGGGCTGGACCGCGCTTGTCTGAGTTGCTGCTTCTGGGTTTCATGAGT 124
 QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 125 CTTGACAAACAGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 184
 QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 185 TCTACAAATGATCTCTACCAAGAACTACAAACACCTAGTACCTTGGAGTACAGCCTG 244
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
 Db 245 CACCTGTGTCTCAACATGCAATGAGCCACAAACATCACAGAAACAGAGTCAAA 304
 QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 305 TTCACATCTACCTGTGTGATAACCTCAGTTTATGGAACACAAACCTCTTGTCCAGTCA 364
 QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 Db 365 CAGACCTCTCTAATCAGCAGAGTGTTCACCAACCCAGCAACGTTTCAACTCCAGAGACA 424
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
 Db 425 ACCTTGAGCCTGACCTGCTGCTGGAATGTTTTCAGACCTTTCACACCTAGCAGTACG 484
 QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
 Db 485 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTCAAGTACATCAAG 544
 QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 Db 545 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATGACTCAGGGCATCTGCTGGAG 604
 QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 Db 605 CAAAATAAGACCTTCAGCTGTGGAGTTTAAAGACAGAGGGAGGGCCCTGGCCCGA 664
 QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 Db 665 GTGCTGTGGGAGGAGCANGCTGATGCTGATGCTGGGCGCCAAAGTATTGCTCTGCTC 724
 QY 221 LeuAla 222
 Db 725 CTTGCC 730

RESULT 6
 AU121093
 LOCUS AU121093 HEMBB1 Homo sapiens cDNA clone HEMBB1002065 5', mRNA
 DEFINITION

sequence.
 ACCESSION AU121093
 VERSION AU121093.1 GI:10936328
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 749)
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isogai,T.

TITLE

HRI human cDNA project

JOURNAL
COMMENT

Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES
source

Location/Qualifiers
 1..749
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HEMBB1002065"
 /clone_lib="HEMBB1"
 /tissue_type="whole embryo, mainly body"
 /dev_stage="embryo, 10 weeks"
 /note="Vector: pME18SFL3"
 BASE COUNT 195 a 211 c 173 g 166 t 4 others
 ORIGIN

Alignment Scores:

Pred. No.: 1..2e-87 Length: 749
 Score: 1078.00 Matches: 214
 Percent Similarity: 96.83% Conservative: 0
 Best Local Similarity: 96.83% Mismatches: 7
 Query Match: 67.76% Indels: 0
 DB: Gaps: 0

US-09-836-602-4 (1-316) x AU121093 (1-749)

QY 1 MetProArgGlyThrAlaLeuSerLeuSerLeuProSerGlyPheMetSer 20

Db 73 ATGCCGGGGCTGGACCGCGCTTGTCTGAGTTGCTGCTTCTGGGTTTCATGAGT 132

QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

Db 133 CTTGACAAACAGGTACTGCTTACCCAGAGTACTACCCAGGGAACATTTTCAATGTT 192

QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60

Db 193 TCTACAAATGATCTCTACCAAGAACTACAAACCTAGTACCTTGGAGTACAGCCTG 252

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80

Db 253 CACCTCTGTCTCAACATGGAATGCAATGAGGCCACAAACATCACAGAAACAGAGTCAAA 312

QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100

Db 313 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCAGTCA 372

QY 101 GlnThrSerValIleSerThrThrValPheThrThrProAlaAsnValSerThrProGluThr 120

Db 373 CAGACCTCTGTAATCAGCAGACAGTGTTCACCAACCCAGCAACGTTTCAACTCCAGAGACA 432

QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140

Db 433 ACCTTGAAGCCTAGCCGCTACCTTGAAATGTTTTCAGACCTTTCACCACTAGCAGTACG 492

QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160

Db 493 CTTGCAACATCTCCACTAAACCTTATACATCATCTTCTCTATCTTCCCTAAGTGACATCAAG 552

QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180

Db 553 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATGACTCANGGCATCTGCTGGAG 612

QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200

Db 613 CAAAATAAGACCTTCAGCTGTGCGAGTTTAAAGAGGACAGGGGAGGCGCTGGCCCGA 672

```

QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 673 GTGCTGTGGGAGGAACANGCTGATGCTGATGCTGGGGCCCAAGTATGCTCTCTCCTCC 732

QY 221 Leu 221
|||
Db 733 TTG 735

RESULT 7
AUI41883          768 bp      mRNA      linear      EST 25-OCT-2000
LOCUS            AUI41883 THYR01 Homo sapiens cDNA clone THYR01001399 5', mRNA
sequence.
ACCESSION        AUI41883
VERSION          AUI41883.1 GI:11003404
KEYWORDS         EST.
SOURCE           human.
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE        1 (bases 1 to 768)
AUTHORS          Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE            HRI human cDNA project
JOURNAL          Unpublished (2000)
COMMENT          Helix Research Institute
                  Genomics Laboratory
                  Contact: Takeo Isogai
                  1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
                  Tel: 81-438-52-3951
                  Fax: 81-438-52-3952
                  Email: genomics@hri.co.jp
                  HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
                  Research Institute; cDNA library construction: Department of
                  Virology, Institute of Medical Science, University of Tokyo, and
                  Helix Research Institute.
FEATURES         Location/Qualifiers
                    source
                      1..768
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="THYR01001399"
                        /clone_lib="THYR01"
                        /tissue.type="thyroid gland"
                        /note="Vector: pME18SFL3"
BASE COUNT      198 a 215 c 181 g 171 t 3 others
ORIGIN
Alignment Scores:
Pred. No.:      7 34e-87      Length:      768
Score:          1069.50      Matches:     217
Percent Similarity: 95.61%      Conservative: 1
Best Local Similarity: 95.18%      Mismatches: 10
Query Match:    67.22%      Indels:    1
DB:              9      Gaps:      0

US-09-836-602-4 (1-316) x AUI41883 (1-768)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 72 ATGCCGCGGGCTGGACCGCTTTGCTGCTGAGTTTGCCTCTCTGGTTTCATGAGT 131

QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 132 CTTGCAACAACAGGACTACTCTACCCACAGAGTTACCTACCAGGAGAACATTTTCAATGTT 191

QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 192 TCTACAAATGATCTCTACCAAGAACTACACACCTAGTACCTTGGAGTAGTACCAGCCTG 251

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
|||||

```

```

Db 252 CACCCCTGTGTCTCAACATGGCAATGAGGCCAACAAACACACACGACGAGTCAA 311
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 312 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 371
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 372 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCAGCGTTTCAACTCCAGAGACA 431
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 432 ACCTTGAACCTAGCCTGTGCACCTGGAAATGTTTCAGACCTTTCAACCACTAGCACTAGC 491
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 492 CTTGCAACATCTCCCACTAAACCTATATACATCATCTTCTCTATCTTAAGTACATCAAG 551
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 552 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCCTGGAG 611
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyLeuAlaArg 200
|||||
Db 612 CAAAATAAGACCTTTCAGCTGTGCGGAGTTTAAAGAGACAGGGGAGGCGCTGGCCCGA 671
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 672 GTGCTGTGGGGAGGAACANGCTGATGCTGATGCTGGGCG--CCAGTATGCTTCTCTGCTC 729
QY 221 LeuAlaGlnSerGluValArgPro 228
|||||
Db 730 CTTGCCAATCTGANGTGAAGGCCT 753

RESULT 8
B1769224
LOCUS            B1769224
DEFINITION       B1769224.1 1129 bp      mRNA      linear      EST 25-SEP-2001
603060051F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5209333 5',
mRNA sequence.
ACCESSION        B1769224
VERSION          B1769224.1 GI:15760802
KEYWORDS         EST.
SOURCE           human.
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE        1 (bases 1 to 1129)
AUTHORS          NIH-MGC http://mgc.nci.nih.gov/.
TITLE            National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL          Unpublished (1999)
COMMENT          Contact: Robert Strausberg, Ph.D.
                  Email: cgapbs-r@mail.nih.gov
                  Tissue Procurement: Life Technologies, Inc.
                  cDNA Library Preparation: Life Technologies, Inc.
                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLAM11525 row: e column: 14
                  High quality sequence stop: 765.
FEATURES         Location/Qualifiers
                    1..1129
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:5209333"
                      /clone_lib="NIH_MGC_122"
                      /lab_host="DH10B"
                      /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                      Site.1: NotI; Site.2: EcoRV (destroyed); RNA source
                      anonymous pool of 24 week female lung, 16 week female
                      spleen, and 20-22 week male spleens. Library is oligo-dT
                      primed and directionally cloned (EcoRV site is destroyed)

```

upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library.*

BASE COUNT 292 a 317 c 279 g 241 t
ORIGIN

Alignment Scores:
Pred. No.: 6,01e-84 Length: 1129
Score: 1040.00 Matches: 247
Percent Similarity: 81.70% Conservative: 12
Best Local Similarity: 77.92% Mismatches: 31
Query Match: 65.37% Indels: 27
DB: 10 Gaps: 7

US-09-836-602-4 (1-316) x B1769224 (1-1129)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
DB 55 ATGCGCGGGGCTGGACCGCGCTTGTGCTGCTGAGTTTCTGCTTCTGGGTTCAATGACT 114
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 115 CTGACAACAACGGTACTCTACCCAGAGTTACCTACCCAGGGACATTTTCAATGTT 174
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 175 TCTACAAATGATCTTACCAAGAACTACACACCTAGTACCTTGGAGTAGTACCGCTG 234
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
DB 235 CACCCTGTCTCAACATGGCAATGAGGCCACAAACAACATCACAGAAACGACAGTCAAA 294
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 295 TTCACATCTACCTCTGTGATACTCAGTTTATGGAAACACAAACCTTCTGTCTCAGTCA 354
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 355 CAGACCTCTGTATACGACAGTGTTCACACCCAGCCCAACGTTTCACTCCAGAGACA 414
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
DB 415 ACCTTGAAGCCTAGCTGTACCTGGAATGTTTTCAGACCTTTCAACCACTAGCAGTACG 474
QY 141 LeuAlaThrSerProThr-LysProTyrThrSerSerProIleLeuSerAspIleY 160
DB 475 CTGGCAACATCTCCCACTAAAGCCCTATACATCATCTTCTCTATCTTAAGTGACATCAA 534
QY 160 sAlaGluIleLysCysSerGlyIle-ArgGluVal-LysLeuThrGlnGlyIleCysLeu 179
DB 535 GGCAGAAATCAATGTTTCAGGATCATCAGAGAGTGAATGACTCAGGCGATCTGCCCTG 594
QY 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
DB 595 GAGCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAAGACAGAGGGGCGCTGGCC 654
QY 200 -ArgValLeuCys--GlyGluGluGlnAlaAspAlaAspAlaGlyAla-Gln--ValC 217
DB 655 CCGAGTGCTGTGCTGGCGGAGGAGGAGGCTGATGCTGTGCTGGGCGCCAGCGATGCG 714
QY 217 ysSerLeuLeuLeuAlaGlnSerGlu-ValArgPro-GlnCysLeuLeuValLeuAl 236
DB 715 TCCCGTGTGCTTGGCCAGCTGACGCTGAGCCCTCCAGTGTCTACTGCTGGGCTTTCG 774
QY 236 a-----AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisG1 252
DB 775 GCCCCTAGTAAGGGGAATTTGCCCGAGCAACAGTCTCCAGCTTTATGAAAAGACCCCA 834
QY 252 nSerAspLeuLysLysLeu-----GlyIleLeuAspPheThrGluGlnAspValAlase 270
DB 835 ATCTGAGCCTGAACAAGCTGGGGGAATCCATAGAAATTTTCGCTGAGCAAGATTGTTTCG 894

QY 270 rHisGlnSerTyrSerGlnLysThrLeuIleAla-----Le 282
DB 895 AAGCATAGC-----AGAAACTATTTCCTCCCAAGAAGACCCCTGATTGGACACT 945
QY 282 uValThr---SerGlyAlaLeuLeuAlaValLeuGly 293
DB 946 GGAGACCCCTCGGAAGCCCTGCTGGATGACTTGGG 982
RESULT 9
BG675028 804 bp mRNA linear EST 01-MAY-2001
LOCUS 602621258F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4746591 5',
DEFINITION mRNA sequence.
ACCESSION BG675028
VERSION BG675028.1 GI:13906424
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Place: LLNL0595 row: d column: 16
High quality sequence stop: 686.
Location/Qualifiers
1. 804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4746591"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 219 a 217 c 191 g 177 t
ORIGIN
Alignment Scores:
Pred. No.: 5,35e-80 Length: 804
Score: 994.00 Matches: 229
Percent Similarity: 91.41% Conservative: 5
Best Local Similarity: 89.45% Mismatches: 10
Query Match: 62.48% Indels: 12
DB: 10 Gaps: 2
US-09-836-602-4 (1-316) x BG675028 (1-804)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
DB 46 ATGCGCGGGGCTGGACCGCGCTTGTGCTGCTGAGTTTGTGCTTCTGGGTTCAATGACT 105
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 106 CTTGACAACAACGGTACTGCTACCCAGAGTTTACCTACCCAGGGAACATTTTCAATGTG 165
QY 41 SerThr-AsnValSerTyrGlnGlu---ThrThrThrProSerThrLeuGlySerThrSe 59
DB 166 TCTACAAAGTGTATCTTACCAAGAAAGTACACAGCACTAGGTACCTTGTGTAAGTACCAC 225
QY 59 rLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVa 79

```
|||||
Db 226 CCTGCACCTGTGTCTCAACATGGCAATGAGGCCACCAACAAACATCAGAAAACGACAGT 285
Qy 79 llySerPheThr-SerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValG 99
Db 286 CAAATTCAGCATCTACCTCTGTGTAACCTCAGTTTATGGAACACAAACATCTTCTGTGCC 345
Qy 99 lnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProG 119
Db 346 AGTCACAGACCTCTGTAATCAGCAGAGTTCACACCCCGACCCACGCTTTCAACTCCAG 405
Qy 119 luThrThr-LeuLysProSerLeuSerProGlyAsnValSerAsp-LeuSerThrThr-S 138
Db 406 AGACAAGCCTTGAAGCCTGAGCCTGTGCACCTGGAATGTTTCAGACCCCTTTCACACCAC 465
Qy 138 erThrSerLeuAlaThrSerProThrLysProThrLysProThrThrSerSerSerProIleLeuSerA 158
Db 466 GCACCTAGCCTTGCACACATCTCCCACTAAACCTATACATCATCTCTCTATCTAAGTG 525
Qy 158 spIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleC 178
Db 526 ACATCAAGCGAGAATCAAAATGTCAGGCATCAGAGAAGTGAATGACTCAGGGCATCT 585
Qy 178 ysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyL 198
Db 586 GCCTGGAGCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAAGACAGGGGAGGGAC 645
Qy 198 euAlaArgValLeuCysGlyGluGluGln-AlaaspAlaasp-AlaGlyAlaGlnVal-- 216
Db 646 TGGCCCGAGTCTGTGTGGGAGGAGCAGAGCTGATGCTGATGCTAGTGGGGGCCAGGTATA 705
Qy 217 -CysSerLeuLeuAlaGlnSer-GluValArgProGlnCysLeuLeu-LeuValLeu 235
Db 706 GCTCCCTGCTACCTGGCCAGCTAGAGGTGAGGCCTCAGCTGCTACTGACTGCTGCTG 765
Qy 236 -AlaAsnArgThrGluIleSerSerLys 244
Db 766 GGCAACAGACAGACATTTTCAGCAAA 793

RESULT 10
LOCUS AU135182 754 bp mRNA linear EST 24-OCT-2000
DEFINITION AU135182 PLACE1 Homo sapiens cDNA clone PLACE1001404 5', mRNA
sequence.
ACCESSION AU135182
VERSION AU135182.1 GI:10995721
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 754)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamanoto,J., Wakamatsu,A., Nakamura,Y., Negai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1..754
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1001404"
```

```
/clone_lib="PLACE1"
/tissue_type="placenta"
/notes="Vector: pME18SFL3"
BASE COUNT 201 a 210 c 172 g 166 t 5 others
ORIGIN
Alignment Scores:
Pred. No.: 8,21e-80 Length: 754
Score: 991.50 Matches: 208
Percent Similarity: 90.60% Conservative: 4
Best Local Similarity: 88.89% Mismatches: 12
Query Match: 62.32% Indels: 10
DB: 9 Gaps: 3
US-09-836-602-4 (1-316) x AU135182 (1-754)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuProSerGlyPheMetSer 20
Db 48 ATGCCGCGGGCTGGACGCGCTTGTGCTGTGAGTTTGTGCTGCTTCTGGGTTCATGAGT 107
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 108 CTTGACAAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGGAACATTTTCAATATT 167
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 168 TCTACAAATGTTATCTACCAAGAAACTACACACCTAGTACCTTGGGAAGTACCAGCCTG 227
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 228 CACCGTGTGTCTCAACATGCGCAATGAGGCCACAACAAACATCACAGAAACGACAGTCAA 287
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 288 TTCACATCTACCTCTGTGATAACTTCAGTTATGGAACACAAACACTCTCTGTCCAGTCA 347
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 348 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCACGTTTCAACTCCAGAGACA 407
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 408 ACCTTGAAGCCTAGCCTGCTACCTCGGAATGTTTCAGACCTTTCAACCACTAGCAGTASC 467
Qy 141 LeuAlaThrSerProThrLysProThrLysProThrSerSerProIleLeuSerAspIleLys 160
Db 468 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCCTAAGTGACATCAAG 527
Qy 161 Ala-GluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuG1 180
Db 528 GCAAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGSCATCTGCCTGGA 587
Qy 180 uGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu---GlyLeuAl 199
Db 588 GCAAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAAGACAAAGGGGAAAGGGCCTGGGC 647
Qy 199 aArgVal---LeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSe 218
Db 648 CCGAGTTGCTGTGTGGGAGGAGCAGAGCTGATGCC-----TG 686
Qy 218 rLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeu 231
Db 687 ATTGCTGGGGGCCCAAAAGTTANNGCTCCCTGNCCTCCTT 726
RESULT 11
Bg256232
LOCUS Bg256232 1001 bp mRNA linear EST 13-FEB-2001
DEFINITION Bg256232 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:447318 5',
mRNA sequence.
ACCESSION Bg256232
VERSION Bg256232.1 GI:12766048
KEYWORDS EST.
SOURCE human.
```



```

Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 274 CTTGACAAACAGGTACTGCTACCCAGAGTACTACCCAGGAACATTTTCAAAATGTT 333
Qy 41 SerThrAsnValSerTyrglnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 334 TCTACAAATGTATCTACCAAGAAACTACACACCTAGTACCCCTGGAGGTACAGCCTG 393
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValValys 80
Db 394 CACCCTGTGTCTCAACATGGCAATGAGGCCACAACAACATCACAGAAACGACAGTCAAA 453
Qy 81 PheThrSerThrSerValIleThrSerValTyrglnGlyAsnThrAsnSerSerValGlnSer 100
Db 454 TTCACATCTACTCTGTGATACCTCAGTTATGGAACACAAACTCTCTGTGTCAGTGCA 513
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 514 CAGACCTCTGTAATCAGACACAGTGTTCACCACCCAGCCAGCGTTTCAACTCCAGAGACA 573
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 574 ACCTTGAAGCCTACCCCTGTACCTGGAAATGTTTCAGACCTTTTCAACCACTAGCAGTAC 633
Qy 141 LeuAlaThrSerProThrLysProTyrglnGlySerSerProIleLeuSerAspIleLys 160
Db 634 CTTGACACATCTTCCATCAACCCATATCA-TCATCTTCTNCTATNCTAAGTGACATCAAG 692
Qy 161 AlaGluLeuLysCysSerGlyIle 168
Db 693 GCAGAAATCAAAATGTTTCANGCATC 716

RESULT 13
AUI41125 733 bp mRNA linear EST 25-OCT-2000
LOCUS AUI41125 THYR01 Homo sapiens cDNA clone THYR0100063 5', mRNA
DEFINITION sequence.
ACCESSION AUI41125
VERSION AUI41125
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 733)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
SOURCE Location/Qualifiers
1..733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR0100063"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/notes="Vector: pME18SPL3"
BASE COUNT 199 a 212 c 168 g 150 t 4 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1,81e-58 Length: 733
Score: 755.50 Matches: 157
Percent Similarity: 94.61% Conservative: 1
Best Local Similarity: 94.01% Mismatches: 8
Query Match: 47.49% Indels: 2
DB: 9 Gaps: 1

US-09-836-602-4 (1-316) x AUI41125 (1-733)

Qy 1 MetProArgGlyThrPheAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 233 ATGCCCGGGGCTGGACCGCGCTTTGCTTGCTGAGTTTGCTGCTTCTGGGTTTCATGAGT 292
Qy 21 LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 293 CTTGACAAACAGGTACTGCTACCCAGAGTTACTACCCAGGAACATTTTCAAAATGTT 352
Qy 41 SerThrAsnValSerTyrglnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 353 TCACAAATGTATCTACCAAGAAACTACACACCTAGTACCTTGGAGGTACCGAGCTG 412
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValValys 80
Db 413 CACCCTGTGTCTCAACATGGCAATGAGGCCACAACAACATCACAGAAACGACAGTCAAA 472
Qy 81 PheThrSerThrSerValIleThrSerValTyrglnGlyAsnThrAsnSerSerValGlnSer 100
Db 473 TTCACATCTACTCTGTGATACCTCAGTTATGGAACACAAACTCTCTGTGTCAGTGCA 532
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 533 CAGACCTCTGTAATCAGACACAGTGTTCACCACCCAGCCAGCGTTTCAACTCCAGAGACA 592
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 593 ACCTTGAAGCCTACCCCTGTACCTGGAAATGTTTCAGACCTTTTCAACCACTAGCAGTAC 652
Qy 141 LeuAlaThrSerProThrLysProTyrglnGlySerSerProIleLeuSerAspIleLys 160
Db 653 TGCACATCTGCC---ACTAACCCCTATCA-TCATCTTCTNCTATCTTCCTAAGTGACATCAAG 708

Qy 161 AlaGluLeuLysCysSerGly 167
Db 709 GCAGAAATCAAAATGTTTCAGCG 729

RESULT 14
AUI39201 825 bp mRNA linear EST 25-OCT-2000
LOCUS AUI39201 PLACE1 Homo sapiens cDNA clone PLACE1010139 5', mRNA
DEFINITION sequence.
ACCESSION AUI39201
VERSION AUI39201
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 825)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of

```

Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

Location/Qualifiers
1 825

```
1. .825
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE10139"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="vector : pME18SLF3"
243 c 185 t
5 others
```

BASE COUNT	227 a	243 c	185 g	165 t	5 others
BASE COUNT	227 a	243 c	185 g	165 t	5 others

Alignment Scores:		
Pred. No.:	1,238-54	825
Score:	714.00	Matches: 148
Percent Similarity:	95.54%	Conservative: 2
Best Local Similarity:	94.27%	Mismatches: 4
Query Match:	44.88%	Indels: 3
DB:	9	Gaps: 0

US-09-836-602-4 (1-316) x AU139201 (1-825)

Qy	1	Met	Pro	Arg	Gly	Thr	Pro	Ala	Leu	Cys	Leu	Leu	Ser	Leu	Pro	Ser	Gly	Phe	Met	Ser	20	
Db	301	AT	CG	CG	GG	GG	CT	GG	AC	CG	CG	CT	TG	CT	CG	AG	TT	TG	CG	CT	360	
Qy	21	Leu	Asp	Asn	Asn	Gly	Thr	Ala	Thr	Pro	Glu	Leu	Pro	Thr	Gln	Gly	Thr	Phe	Ser	Asn	Val	40
Db	361	CT	TG	GAC	AAC	AAC	GGT	ACT	GT	CT	ACC	CGA	GTT	ACT	TAC	CAG	GGG	AAC	ATT	TTC	AAAT	420

Qy	41	SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu	60
Db	421	TCACAAATGTCTCCACCAAGAACTACAAACCTAGTACCTTGGAGTAGTACACGCTG	480
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
Db	481	CACCCCTGTCTCAACATGCGAATGAGGCCAACAAACATCACAGAAACGACAGTCAA	540

Qy	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerValGlnSer	100
Db	541	TTCACTCTACCTCTGTGATAACCTCAGTTATATGGAAACACAAACTTCTGTCTCACTCA	600
Qy	101	Gln-ThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluTh	120
Db	501	CAGACCTCTGTATCAGCACAGTGTTCANCAACCCAGCCACCGTTTCAATCCAGAGAC	660

Qy	120	rThrLeuLysProSerLeuSerProGly-AsnValSerAspLeuSerThrSerThr-S	140
	:.....:.....:.....:.....:.....:.....:.....:.....:.....	
Db	661	AACCTTGAAGCCCTAACCTGTCACCTGSGAAATGCTTCAAACTTTTAAACCACTTANCACTA	720
Qy	140	erLeuAlaThrSerProThr-LysProTyrThrSerSerPro	154
Db	721	GCCTTGGACACATCTCCCACTAAACCCATATACATCATCTTCTCT	765

RESULT 15					
B1824944					
LOCUS	B1824944	903 bp	mRNA	linear	EST 04-OCT-2004
DEFINITION	G03033307f1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174593				
	mRNA sequence.				
ACCESSION	B1824944				
VERSION	B1824944.1				
KEYWORDS	EST				GI:15936494

SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 903) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-re@mail.nih.gov
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

```
ncpp://image.llnl.gov
Plate: LLAM11434 row: n column: 02
High quality sequence start: 6
High quality sequence stop: 587.
```

FEATURES source

```

1. .903
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5174593"
/clone_lib="NIH_MGC_ll5"
/lab_host="DH10B"

```

/note="Organ: pooled brain, lung, testis; Vector: pCMV-Sport6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."

BASE COUNT	291 a	262 c	203 g	147 t
------------	-------	-------	-------	-------

Alignment Scores:

Pred. No.:	5,32e-52	Length:	903
Score:	685.50	Matches:	165
Percent Similarity:	81.43%	Conservative:	6
Best Local Similarity:	78.57%	Mismatches:	26
Query Match:	43.09%	Indels:	13
DB:	10	Gaps:	2

US-09-836-602-4 (1-316) x BI824944 (1-903)

Qy	1	MetProArgGlyTrpThrAlaLeuCys - LeuLeu - SerLeuLeuProSerGlyPheMet-	19
Db	99	ATGCCGGGGCGCTGGACCGGCGCTTTGCTATGCTGAGATTGCTGCTCTGGGTTTCATGA	158
Qy	20	SerLeuAspAsnAsnGly - ThrAlaThrProGluLeuLeuProThrGlnGlyThrPheSerAs	39
Db	159	GATCTTGACAAACAGGATACTGCTACCCAGAGATTACCTACCCAGGGGAACATTTCAA	218
Qy	39	nValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSe	59
Db	219	TGTTTCTACAAATGTATCTTCACCAAGAAACTACAAACACCTAGTACCTTGGAAAGTACCAG	278
Qy	59	rLeuHisProValSerGlnHisGlyAsnGluAlaThr - ThrAsnIleThrGluThrThrV	79
Db	279	CTGCAACCTTGTCCTCACTGGCAATGAGGCCAACAACAAACATCAGAAACGACAG	338
Qy	79	allYsPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValc	99
Db	339	TCAAAATTCACATCTACCTCTGTGATAAACCCTCAGTTTATGGAAACACAAACTCTCTGTCC	398
Qy	99	lnSerGlnThrSerValIleSerThrValPheThrThrPro - AlaAsnValSerThrPro	118
Db	399	AGTCACAGACCTGTGTAATCAGCACAGTGTTCACCAACCCAGACCAACGTTTCAACTCCA	458
Qy	119	GluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThr - - -	137
Db	459	GAGACAACTTGAAGCCTAGCCTGCTCACCTGGAAATGTGTCAGACCTTTGCAAAACACTAG	518
Qy	138	SerThrSerLeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSer	157
Db	519	CACATAGCCATGACAAATCCTCCCACTAAACCCCTATACATCATCTTCTCCTATCCTAAGT	578
Qy	158	AspIleLysAla - GluIleLysCysSerGlv - IleArg - - - - - GluValLysLeuThr	174

```
||||| 579 GACATCAGGCCAGAACTCAAAATGTCGGCAATCAGACAAAGCGAAACTGACTCCGGG 638
QY 175 GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGlu-PheLysLysAspAr 194
Db 639 CCATCGGCCAAGGCAGCAAAACCTAAAGACCTCCAGACTGGCGGGAAGCACACCGAAGGACAA 698
QY 194 gGlyGluGly 197
Db 699 GGGCACAGGG 708
```

Search completed: October 30, 2002, 10:47:56
Job time : 1100.74 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 30, 2002, 08:08:13 ; Search time 40.5374 Seconds
(without alignments)
1914.781 Million cell updates/sec

Title: US-09-836-602-4
Perfect score: 1591
Sequence: 1 MPRGWTALCLSLPSGEMS.....YFLMNRSSPTGERLELEP 316

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlip
-Q=/cgn2_1/USPTO.spool/US09836602/runat_29102002_101145_3837/app_query.fasta_1.1429
-DB-issued_Patents_NA -Qfmt=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=-1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602_8CGN1_1_44_@runat_29102002_101145_3837 -NCFU=6 -ICPU=3
-NO_XLPAY -NO_MMWP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	921	57.9	1065	2 US-08-475-634D-18
2	144	9.1	5163	3 US-08-700-651-1
3	144	9.1	5163	3 US-08-928-361B-4
4	144	9.1	5318	3 US-08-700-651-2
5	144	9.1	5318	3 US-08-928-361B-3
6	137.5	8.6	5511	3 US-08-928-361B-2
7	137.5	8.6	7334	3 US-08-928-361B-1
8	129	8.1	1505	1 US-07-915-246-1
c 9	128	8.0	390	4 US-09-137-649-7
c 10	127.5	8.0	3337	1 US-08-072-610-1
c 11	127.5	8.0	3337	2 US-08-719-822B-1
c 12	127.5	8.0	3337	4 US-09-092-458-1

13	125.5	7.9	2589	1	US-08-325-267A-3	Sequence 3, Appli
14	125.5	7.9	2685	3	US-08-362-525-21	Sequence 21, Appl
15	125.5	7.9	4614	1	US-08-325-267A-1	Sequence 1, Appli
16	123	7.7	2754	1	US-08-270-076A-10	Sequence 10, Appl
17	121	7.6	2214	6	5258502-1	Patent No. 5258502
18	121	7.6	3168	4	US-09-165-239A-3	Sequence 3, Appli
19	119.5	7.5	688	4	US-08-998-416-915	Sequence 915, App
20	119.5	7.5	1107	2	US-08-991-300-1	Sequence 1, Appli
21	117	7.4	2584	3	US-08-758-662-8	Sequence 8, Appli
c 22	116	7.3	2793	1	US-08-209-747-1	Sequence 1, Appli
c 23	116	7.3	2793	1	US-08-458-298-1	Sequence 1, Appli
24	115.5	7.3	2032	4	US-09-241-581B-5	Sequence 5, Appli
25	115.5	7.3	2032	5	PCT-US95-07721-5	Sequence 5, Appli
26	114.5	7.2	2093	1	US-08-287-001A-1	Sequence 1, Appli
27	114.5	7.2	2093	5	PCT-US95-09941-1	Sequence 1, Appli
28	113.5	7.1	2164	4	US-08-760-615-3	Sequence 3, Appli
29	113	7.1	36519	3	US-08-923-137-2	Sequence 2, Appli
30	108.5	6.8	4108	4	US-08-981-729-8	Sequence 8, Appli
31	108.5	6.8	4108	4	US-08-981-446B-1	Sequence 1, Appli
32	108	6.8	2150	2	US-08-861-464-13	Sequence 13, Appl
33	108	6.8	2150	2	US-08-396-001-13	Sequence 13, Appl
c 34	108	6.8	2150	4	US-08-323-433A-13	Sequence 13, Appl
c 35	108	6.8	9636	1	US-08-323-170B-1	Sequence 1, Appli
36	108	6.8	9636	4	US-08-954-441-1	Sequence 1, Appli
37	107.5	6.8	3833	1	US-08-917-320-18	Sequence 18, Appl
38	107.5	6.8	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
39	107.5	6.8	5931	3	US-08-783-774-1	Sequence 1, Appli
40	106	6.7	10803	3	US-09-080-044-1	Sequence 1, Appli
41	106	6.7	19227	3	US-09-090-793-13	Sequence 13, Appl
42	106	6.7	40138	3	US-09-090-793-12	Sequence 12, Appl
43	105	6.6	3141	2	US-08-658-665-66	Sequence 66, Appl
44	105	6.6	3141	4	US-08-796-101-30	Sequence 30, Appl
45	105	6.6	3141	4	US-09-085-273-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-08-475-634D-18
; Sequence 18, Application US/08475634D
; Patent No. 5962844
; GENERAL INFORMATION:
; APPLICANT: HAWLEY, Robert J.
; APPLICANT: MONROY, Rodney L.
; TITLE OF INVENTION: Antibodies to Porcine CD34 Positive
; TITLE OF INVENTION: and Uses Therefor
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS 6.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,634D
; FILING DATE: June 7, 1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 61750-141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 973-994-1700
; TELEFAX: 973-994-1744
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:

Qy 141 uAlaThrSerProThrLysProTyrThrSerSerSer 153
 Db 973 CACCACAAACAACAACAACAACAACAACAACAACAACA 1009

RESULT 5

US-08-928-361B-3
 ; Sequence 3, Application US/08928361B
 ; Patent No. 6071518
 ; GENERAL INFORMATION:
 ; APPLICANT: Petersen, Carolyn
 ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
 ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
 ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
 ; TITLE OF INVENTION: SPECIES INFECTIONS
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
 ; STREET: 385 Sherman Avenue, Suite 6
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306-1840
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/928,361B
 ; FILING DATE: 12-SEP-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/026,062
 ; FILING DATE: 13-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Verny, Hana
 ; REGISTRATION NUMBER: 30,518
 ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
 ; TELEPHONE: 650-324-1677
 ; TELEFAX: 650-324-1678
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5318 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-928-361B-3

Alignment Scores:
 Pred. No.: 7.04e-05 Length: 5318
 Score: 144.00 Matches: 37
 Percent Similarity: 47.37% Conservative: 26
 Best Local Similarity: 27.82% Mismatches: 69
 Query Match: 9.05% Indels: 1
 DB: 3 Gaps: 0

US-09-836-602-4 (1-316) x US-08-928-361B-3 (1-5318)

Qy 22 AspAsnAsnGly-ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSe 41
 Db 612 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 671
 Qy 41 rThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeuHi 61
 Db 672 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 731
 Qy 61 sProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLysPh 81
 Db 732 ACCGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 791
 Qy 81 eThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGl 101

Db 792 TACAACCACTACAAACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 851
 Qy 101 nThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrTh 121
 Db 852 AACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 911
 Qy 121 rLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrThrSerLe 141
 Db 912 CAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 971
 Qy 141 uAlaThrSerProThrLysProTyrThrSerSerSer 153
 Db 972 CACCACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1008

RESULT 6

US-08-928-361B-2
 ; Sequence 2, Application US/08928361B
 ; Patent No. 6071518
 ; GENERAL INFORMATION:
 ; APPLICANT: Petersen, Carolyn
 ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
 ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
 ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
 ; TITLE OF INVENTION: SPECIES INFECTIONS
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
 ; STREET: 385 Sherman Avenue, Suite 6
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306-1840
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/928,361B
 ; FILING DATE: 12-SEP-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/026,062
 ; FILING DATE: 13-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Verny, Hana
 ; REGISTRATION NUMBER: 30,518
 ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
 ; TELEPHONE: 650-324-1677
 ; TELEFAX: 650-324-1678
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5511 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-928-361B-2

Alignment Scores:
 Pred. No.: 0.000377 Length: 5511
 Score: 137.50 Matches: 61
 Percent Similarity: 39.19% Conservative: 46
 Best Local Similarity: 22.34% Mismatches: 135
 Query Match: 8.64% Indels: 31
 DB: 3 Gaps: 7

US-09-836-602-4 (1-316) x US-08-928-361B-2 (1-5511)

Qy 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
 Db 972 CACCACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1008


```
Qy 204 GlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGln 223
Db 3334 ATCGAAAGATGACGCTATTGAAATGGAATGCATTTACAATGATCCAAATGATGAC 3393
Qy 224 SerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGluIle----- 241
Db 3394 ACGCATGTCCTTCAGATTTAAGGTTAAAGATGATAGGAACACTATTTCAGTAAGATGC 3453
Qy 242 -----SerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeu 258
Db 3454 GGAAGAGTGCGAGTAACCTCGAGTTCACAGATAGAGT----- 3492
Qy 259 GlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThr 278
Db 3493 -----TTGGATTTCACAAATTCCTCCAGTAGCTGGCCATACACTGTTC----- 3537
Qy 279 LeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaVal 291
Db 3538 ATAATAGTTGTGTGAGCGCGGTGGAAAAATTCACGTA 3576
RESULT 8
US-07-915-246-1/c
; Sequence 1, Application US/07915246
; Patent No. 5401836
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Chris L.
; APPLICANT: Fallis, Lynne
; APPLICANT: Bellmare, Guy
; APPLICANT: Bolvin, Rodolphe
; TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
; TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held, and Malloy
; STREET: 500 W. Madison St. 34th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/915,246
; FILING DATE: 19920716
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochoplen, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 91 P 1125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 707-8889
; TELEFAX: 312 707-9155
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1505 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; STRAIN: Westar
; DEVELOPMENTAL STAGE: Somatic
; TISSUE TYPE: Root
US-07-915-246-1
```

Alignment Scores:

Pred. No.: 0.000333 Length: 1505

```
Score: 129.00 Matches: 47
Percent Similarity: 48.78% Conservative: 33
Best Local Similarity: 28.66% Mismatches: 70
Query Match: 8.11% Indels: 14
DB: 1 Gaps: 5
US-09-836-602-4 (1-316) x US-07-915-246-1 (1-1505)
Qy 12 SerLeuLeuProSerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeu 31
Db 1370 AGTGCTATACCTCCCTCCTCCTCCGCCACCTCCACACCCACCGCCCTCCTCC 1311
Qy 32 ProThrGlnGlyThrPheSer-----AsnValSerThrAsnValSerTyr 46
Db 1310 ACCACCGTGCTCTCTCCACCCACCCACCGATATCCCTCTCCAGCTCCACCTCCGGC 1251
Qy 47 GlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSerGlnHis 66
Db 1250 ACCATAACCAACCGGTGAGCTCCACGGGGCCAGAACCTCCTCCTCCACCGCTCCTTT 1191
Qy 67 GlyAsnGluAlaThrThrAsnIle-----ThrGluThrThrValLysPheThrSer 83
Db 1190 TCCCCCGCTCCACCCACCTCCATGTCCACCTGACCGCCACCTCCATATCCTCCTCCAGC 1131
Qy 84 ThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGlnThrSer 103
Db 1130 ACCAGCGCTTCTCCACCCACCGTATCCACCCGCTGTCACCTCCACCTCCTCCACCGCC 1071
Qy 104 ValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrThrLeuLys 123
Db 1070 TCCATTTCCACCGCTCCACCGACC-----TCCATGTCCACCTGCACCGCCACC----- 1023
Qy 124 ProSerLeuSerProGlyAsnValSerAspLeuSerThrThr-----SerThrSerLeuAla 142
Db 1022 ---TCCATATCCTCCTCCACCGACCGACCTTCTCCACCGCTATCCACCGATCCACCGCCG 966
Qy 143 ThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLysAlaGlu 162
Db 965 ACCGCCAGCTCCTCCACCGACCGCTCCTCGCTTCTCCTCCACCTCCACCATGTCTCCTCCAGC 906
Qy 163 IleLysCysSer 166
Db 905 TCCTCCGCTCC 894
RESULT 9
US-09-197-649-7
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Pribnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/CI-CON
; CURRENT APPLICATION NUMBER: US/09/197,649
; CURRENT FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: 07/829,461
; EARLIER FILING DATE: 1992-01-31
; EARLIER APPLICATION NUMBER: 07/739,055
; EARLIER FILING DATE: 1991-08-01
; EARLIER APPLICATION NUMBER: 07/561,968
; EARLIER FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.
```



```
DB: 4 Gaps: 6
US-09-836-602-4 (1-316) x US-09-092-458-1 (1-3337)
QY 5 TrpThrAlaLeuCysLeuSerLeuLeuPro-----SerGlyPheMetSerLeu 21
DB 2527 TGGTAGTCATGTCGATGATAGATAAAATCCAGACGCTAAATCGCTTCTTTGGTTG 2468
QY 22 AsnAsn-----AsnGlyThrAlaThrProGluLeuProThrClnGlyThr 36
DB 2467 TCTAATGTATATCTCTCCGATGGTATTAATAATGTAATACATCTCTTCCTCTCT 2408
QY 37 PheSerAsnValSerThrAsnValSerThrGlnGluThrProSerThrLeuGly 56
DB 2407 GGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2348
QY 57 SerThrLeuHisProValSerGlnHisGlyAsnGluAlaThrAlaThrAsnIleThrGlu 76
DB 2347 GGTACTTCTACTTCTACTACCGCTGGTACCTCTACTTCTCTACTTCTTCCACTCTTCC 2288
QY 77 ThrThrValLysPheThrSerThrSerValIleThrSerValThrGlyAsnThrAsnSer 96
DB 2287 ACTTCTCTGGTACCTCTTCCACTTCTCTCTGGTACCTCTTCCACTCTCTCTGGTACCTCT 2228
QY 97 SerValGlnSerGlnThrSerValIleSerThrValPheThrProAlaAsnValSer 116
DB 2227 TCCACTTCTCTGGTACCTCT-----TCCACTTCTCTACTTCTGCGGGTACCTCTTCT 2174
QY 117 ThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThr 136
DB 2173 ACTTCTCTGGTACCTCTTCCACTTCTCTACTTCTGCGAGGTACCTCT-----TCCACT 2120
QY 137 ThrSerThrSerLeuAlaThrSerProThr---LysProThrThrSerSerSerProIle 155
DB 2119 TCTTCTACTTCTGCGAGGTACCTCTTCCACTTCTCTCTCTCTGCGAGGTCTTCTTCCGCC--- 2063
QY 156 LeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGln 175
DB 2062 ---TCTACGGTTCCTCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2006
QY 176 GlyIleCysLeuGluGlnAsnLysThrSerSer 186
DB 2005 GGAGTTGGCTCTAAATCTCTTCTTCTGCGACTTCT 1973
RESULT 13
US-08-325-267A-3
; Sequence 3, Application US/08325267A
; Patent No. 5385271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; TITLE OF INVENTION: CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2589 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: ABXL-ID
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2589
US-08-325-267A-3
Alignment Scores:
Pred. No.: 0.00203 Length: 2589
Score: 125.50 Matches: 85
Percent Similarity: 41.32% Conservative: 53
Best Local Similarity: 25.45% Mismatches: 109
Query Match: 7.89% Indels: 89
DB: 1 Gaps: 16
US-09-836-602-4 (1-316) x US-08-325-267A-3 (1-2589)
QY 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
DB 1099 ACAACACGACACTGAA---CCATGGAGCTGGTACTTTTACTTCGACTTCCACTGAAATGTCT 1155
QY 46 -----TyrGlnGluThrThr-----ThrPro 52
DB 1156 ACTGTCACTGGAAACCAAGCGTTGCCAAGTGAAGAACTGTCATTTGTCATAAACTCCA 1215
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
DB 1216 ACTACTGCCATCTCATCCAGTTTGCA-----TCATCATCTTCAGGACAAATCACCACG 1269
QY 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValThrGly 92
DB 1270 TCTATCAGC-----TCTTCGCGTCCAAATTATTACCCCATTTCTATCTCT 1311
QY 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
DB 1312 AGCAATGGAACCTCTGTGATTTCT---TCTCAGTAATTTCT----- 1350
QY 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
DB 1351 TCTTCAGTCACTTCTCTATTCATCTTCTTCCAGTCACTTCTTCCAGTCACTTCT 1410
QY 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
DB 1411 TCTTCTACAACAACCTCCACTTCTATATTTTCTGAATCATCTAAATCATCCGTCATCTCA 1470
QY 146 ThrLysProThrThrSerSerProIleLeuSerAspIleLysAlaGluIleLysCys 165
DB 1471 ACCAGTAGTTCACACCTCTGGTTCT-----TCTGAGAGCGAAACG 1509
QY 166 SerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSer 185
```


Db 1705 ACATATTTCTTCTTCATTCACCACTTGTACCAGTGCACACACAGCCAGGAATGCT 1764
QY 196 GluGlyLeuAlaArgValLeuCysGlyGluGluGluAlaAspAla----- 210
Db 1765 TCTTCATTAC--CACGTGTACCACTACAAAACAGCAGCAACCACTTTGGTTACCG 1822
QY 211 ---AspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGlnSerGluValArgProGln 229
Db 1823 TGACATCTCGCAGTCTCAATGTGACACGATTCATCCCTGGGATGTTTCCACAG 1882
QY 230 CysLeuLeuValLeuAlaAsnArgThrGluIle---SerSerLysLeuGlnLeuMet 248
Db 1883 CTACTGTTACTGTAGCGCGTCACAAACAGATATACCACATGGTCCCTATTCTTACTA 1942
QY 249 LysLysHisGlnSerAspLeuLysLeuGlyIleLeuAspPheThr----- 264
Db 1943 CAGAGACAAACAGC-----AAACCAA-GGGCAACACAGCAACACACAGAAACAA 1995
QY 265 -----GluGlnAspValAlaSerHisGlnSer 273
Db 1996 AAACAAACCCAGGTAGTACAAATTTCTTCTGTGAATCTGACGTATGCTCT----- 2046
QY 274 TyrSerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGly 293
Db 2047 -----AAGACTGCTTCTCCAGCATTGTATCTACAAGCACTGCTACTATTAAACGC 2097
QY 294 IleThrGlyTyrPheLeuMetAsnArgArgSerTrpSerPro 307
Db 2098 GTTACTACAGATAC-----ACAACATGGTGTCT 2127
RESULT 15
US-08-325-267A-1
; Sequence 1, Application US/08325267A
; Patent No. 585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; TITLE OF INVENTION: CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220

; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: ABXL-ID
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4614
; US-08-325-267A-1
Alignment Scores:
Pred. No.: 0.00549 Length: 4614
Score: 125.50 Matches: 85
Percent Similarity: 41.32% Conservative: 53
Best Local Similarity: 25.45% Mismatches: 109
Query Match: 7.89% Indels: 89
DB: 1 Gaps: 16
US-09-836-602-4 (1-316) x US-08-325-267A-1 (1-4614)
QY 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
Db 3124 ACAACCAACCACTGAA--CCATGGACTGGTACTTTTACTTCGACTTCCCACTGAATGTCT 3180
QY 46 -----TyrGlnGluThr-----ThrPro 52
Db 3181 ACTGTCACCTGGAAACCAATGGCTTGCACACTGATGAACCTGTCATGTTGTCAAACTCCA 3240
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
Db 3241 ACTACTGCCATCTCATCCAGTTGTCA-----TCATCATCTTCAGACACAAATCACCAGC 3294
QY 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValTyrGly 92
Db 3295 TCTATCACG-----TCTTCGCTGCCAATATTATCCCAATTTCTCTCTCAGTCAATTTCT 3336
QY 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
Db 3337 AGCAATGGAACTCTCTGATTCT---TCCTCAGTAATTTCT----- 3375
QY 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
Db 3376 TCCTCAGTCACTTCTCTCTATTCACTTCTTCTCCAGTCAATTTCTCTCTCAGTCAATTTCT 3435
QY 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
Db 3436 TCTTCTACAAACACCTCCACTTCTATATTTCTGAATCATCTAAATCATCCGTCATTTCCA 3495
QY 146 ThrLysProTyrThrSerSerProIleLeuSerAspIleLysAlaGluIleLysCys 165
Db 3496 ACCAGTAGTTCCACCTCTGGTTCT-----TCTGACGCGAAACG 3534
QY 166 SerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlnAsnLysThrSer 185
Db 3535 AGTTCAGCTGGTTCT 3594
QY 186 -----SerCysAlaGluPheLysLysAspArgGly 195
Db 3595 ACATATTTCTTCTCATCATTCACCTTGTATCCAGTGCACACAAAGCCAGGAATGCT 3654
QY 196 GluGlyLeuAlaArgValLeuCysGlyGluGluGluAlaAspAla----- 210
Db 3655 TCTTCTATTAC--CACCTGTCTACCACTACAAAACAGCAGCAACCACTTTGGTTACCG 3712
QY 211 ---AspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGlnSerGluValArgProGln 229
Db 3713 TGACATCTCGCAGTCTCAATGTGACACGATTCATCTCCCTCGGATTTGTTTCCACAG 3772

```
Qy 230 CysLeuLeuValLeuAlaAsnArgThrGluIle---SerSerLysLeuGlnLeuMet 248
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3773 CTACTGTTACTGTTAGCGCGTCACACAGAGTATACCACATGGTGCCTATTCTACTA 3832
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 249 LysLysHisGlnSerAspLeuLysLysLeuGlyIleLeuAspPheThr----- 264
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3833 CAGAGACAACAAGC-----AACCAAA-GGGACAACAGACAACCACAGAAACAACA 3885
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 265 -----GluGlnAspValAlaSerHisGlnSer 273
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3886 AAACAACCACGGTAGTTACAATTCTTGTGTGAATCTGACGTATGCTCT----- 3936
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 274 TyrSerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuAlaValLeuGly 293
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3937 -----AAGACTGCTTCTCCAGCCATTGTATCTACAGCACTGCTACTATTAAACGGC 3987
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 294 IleThrGlyTyrPheLeuMetAsnArgSerTrpSerPro 307
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3988 GTTACTACAGAATAC-----ACAACATGGTGTCTCT 4017
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: October 30, 2002, 10:50:42
Job time : 56.5374 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:04:33 ; Search time 143.956 Seconds
(without alignments)
3768.837 Million cell updates/sec

Title: US-09-836-602-4

Perfect score: 1591

Sequence: 1 MPRGTALCLLSLLPSGFWMS.....YFLMRRNSWSPGTGERLELEP 316

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09836602/runat_29102002_101143_3758/app_query.fasta_1.1429
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602 @CEN_1.1.312 @runat_29102002_101143_3758 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1591	100.0	951	24	Human CD34 surface
2	1574	98.9	1122	24	Human full-length
3	1574	98.9	2453	22	Nucleotide sequenc
4	1574	98.9	2615	21	Human CD-34 polynu
5	1574	98.9	3490	21	Human adenosine re
6	1574	98.9	3490	21	Human CD-34 polynu
7	1574	98.9	3490	21	Human adenosine re
8	1509	94.8	906	24	Human CD34 surface
9	921	57.9	1065	18	Porcine CD34 from
10	508	31.9	615	21	Human secreted pro
11	311	19.5	233	22	Human breast cell
12	311	19.5	233	22	Human foetal liver
13	311	19.5	233	22	Probe #14530 for g
14	311	19.5	233	22	Human brain expres
15	311	19.5	233	22	Human bone marrow
16	311	19.5	233	22	Probe #13951 for g
17	311	19.5	233	22	Probe #18003 used
18	311	19.5	233	22	Probe #9593 used t
19	265	16.7	209	22	Human breast cell
20	265	16.7	209	22	Human foetal liver
21	265	16.7	209	22	Probe #14140 for g
22	265	16.7	209	22	Human brain expres
23	265	16.7	209	22	Human bone marrow
24	265	16.7	209	22	Probe #13530 for g
25	265	16.7	209	22	Probe #17596 used
26	265	16.7	209	22	Probe #9203 used t
27	233	14.6	235	22	Novel human diagno
28	170.5	10.7	406	22	Human breast cell
29	170.5	10.7	406	22	Human foetal liver
30	170.5	10.7	406	22	Probe #4241 for ge
31	170.5	10.7	406	22	Human brain expres
32	170.5	10.7	406	22	Human bone marrow
33	170.5	10.7	406	22	Probe #4329 for ge
34	170.5	10.7	406	22	Probe #4456 used t
35	159	10.0	13154	20	Probe #4210 used t
36	159	10.0	13154	20	Enterococcus faeco
37	150.5	9.5	2336	23	Drosophila melanog
38	148	9.3	40875	18	Insert from cosmid
39	147.5	9.3	7720	21	Genomic DNA encodi
40	146	9.2	4116	23	Drosophila melanog
41	146	9.2	6116	23	Drosophila melanog
42	145.5	9.1	5162	23	Drosophila melanog
43	144.5	9.1	8298	22	Human immune/haem
44	144	9.1	5163	19	Cryptosporidium pa
45	144	9.1	5163	21	ORF encoding a por

ALIGNMENTS

RESULT 1
AAD22662
ID AAD22662 standard; DNA; 951 BP.
XX AAD22662;
AC AAD22662;
XX
XX 26-FEB-2002 (first entry)
DT
DT Human CD34 surface antigen truncated variant (tCD34) encoding DNA.
DE
DE Human; surface marker; surface antigen; T lymphocyte; gene therapy;
KW CD34 truncated variant; tCD34; ds.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH CDS 1..951
FT

```

/*tag= a
/product= "Human CD34 surface antigen
truncated variant (tCD34)"

```

EP1148066-A1.

24-OCT-2001.

18-APR-2001: 2001EP-0109374

18-APR-2000; 2000DE-1019075.

(ZAND/) ZANDER A R.

Zander AR;

WPI; 2002-019289/03.

P-PSDB; AAE13542.

New gene transfer vector (accession number D5M13396) containing a transgene and a nucleic acid sequence coding for a surface marker, useful in gene therapy, and for detecting genetically modified cells or cells which do not express CD34.

Claim 3; Page 12-13; 28pp; English.

The patent discloses a gene transfer vector (accession number DSM13396) containing a transgene and a nucleic acid sequence coding for a surface marker. The surface marker is the CD34 surface antigen, its fragment or variant. The vector is useful for in vitro transduction of T lymphocytes, for gene therapy, and in the enrichment, detection and analysis of cells in vitro that do not naturally express CD34. T lymphocytes transduced with the vector are also useful in gene therapy. The CD34 nucleic acid sequences (marker genes), their fragments or variants are used for detecting genetically modified cells or cells which do not naturally express CD34. The present sequence is a DNA encoding human CD34 surface antigen truncated variant (tCD34).

SQ Sequence 951 BP; 261 A; 264 C; 213 G; 213 T; 0 other;

Alignment Scores:

Pred. No.:	6,44e-117	Length:	951
Score:	1591.00	Matches:	316
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-09-836-602-4 (1-316) x AAD22662 (1-951)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20

Db
1 ATGCCGGGGGCTGGACCGGCTTTGCTGAGTTTGCTGCCCTTCTGGGTTTCATGAGT 60

QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

Db 61 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTT 120

Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60

Db 121 TCTACAAATGTATCCTACCAAGAACTACAACACCTAGTACCCCTTGGAAGTACCAAGCCTG 180

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80

Db 181 CACCCTGTCTCAACATGGCAATGAGGCCACAACAATCACAGAAACGACAGTCAA 240

Qy 81 pheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100

Db 241 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAAACACAACTCTTCTGTCCAGTCA 300

Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120

Db 301 CAGACCTCTGTAATCAGCCACAGTGTTCACCACCCAGCCAACTTCAACTCCAGAGACA 360

Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	361	ACCTTTGAAGCCCTAGCCTGTCACCTGGAAATGTTTTCAGACCTTTTCACCACTAGCACTAGC	420
Qy	141	LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
Db	421	CTTGGCAACATCTCCCACTAAACCCCTATACATCATCTTCTCCTATCTCTAAGTGACATCAAG	480
Qy	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	481	CGAGAAATCAATGTTCAGGCATCAGAGAGAGTGAATTTGACTCAGGGCATCTGCCTGGAG	540
Qy	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	541	CAAAATAAGACCTCCAGCTGTGCGGAGCTTTTAAGAAGGACAGGGGAGGGCCTGGCCCGA	600
Qy	201	ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	601	GTGCTGTGTGGGGAGGAGCAGGCTGATGCTGCTGGGCCCAGGTATGCTCCCTGCTC	660
Qy	221	LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu	240
Db	661	CTTGCCCACTGAGGTGAGGCCTCAGTGTCTACTGCTGGTCTTGCCCAACAGAACAGAA	720
Qy	241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle	260
Db	721	ATTTCCAGCAACTCCAACTTATGAAAAGCACCAATCTGACCTGAAAAAGCTGGGGATC	780
Qy	261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle	280
Db	781	CTAGATTCTACTGAGCAAGATGTTGCAAGCCACCAAGCTATTTCCTCCAAAGACCCCTGATT	840
Qy	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet	300
Db	841	GCATCGTGCACCTCGGGAGCCCTGCTGGCTGCTCTTGGGCATCACTGGCTATTTCCTGATG	900
Qy	301	AsnArgArgSerTrpSerProThrGlyGluArgLeuLeuGluPro	316
Db	901	AATCGCGCAGCTGGAGCCCCACAGAGAAAGGCTGGAACCTAGAACCA	948
RESULT	2		
AAD22661			
ID	AAD22661	standard; DNA; 1122 BP.	
XX			
AC	AAD22661;		
XX			
DT	26-FEB-2002	(first entry)	
XX			
DE	Human full-length CD34 (f1CD34)	surface antigen encoding DNA.	
XX			
KW	Human; surface marker; surface antigen; T lymphocyte; gene therapy;		
KW	full-length CD34; f1CD34; ds.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1122	
FT		/*tag= a	
FT		/product= "Human full-length CD34 surface antigen"	
XX			
PN	EP1148066-A1.		
XX			
PD	24-OCT-2001.		
XX			
PF	18-APR-2001; 2001EP-0109374.		
XX			
PR	18-APR-2000; 2000DE-1019075.		
XX			
PA	(ZAND/) ZANDER A R.		
XX	Zander AR;		
XX			

XX SQ Sequence 2463 BP; 589 A; 709 C; 580 G; 585 T; 0 other;

Alignment Scores:
Pred. NO.: 4,89e-115 Length: 2463
Score: 1574.00 Matches: 313
Percent Similarity: 99.37% Conservativity: 1
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 98.93% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-4 (1-316) x AAH75132 (1-2463)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCCGGCGGGCTGGACCGCGCTTGGTGTGCTGAGTTGCTGCTTCTGGGGTTCATGAGT 186

Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACAAACACGGTACTGCTACCCACAGTTACCTACCCAGGGAACATTTTCAAAATGTT 246

Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTCAAAATGATCTCTACCAAGAACTACAACACCTAGTACCCCTTGGGAAGTACCAGCGCTG 306

Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 307 CACCCTGTGTCTCAACATGGCAATGAGCCACACAAACATCACAGAACACAGTCAAA 366

Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATACCTCAGTTTATGGAACACAAAACCTCTTCTGTCCAGTCA 426

Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCAGCCCAACGTTTCAACTCCAGAGACA 486

Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 487 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTCAGACCTTTCAACCACCTAGCACTAGC 546

Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 547 CTTGCAACATCTCCACTAAACCCCTATACATCATCTCTCTCTATCCTAAGTGACATCAAG 606

Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 607 GCAGAAATCAAAATGTTACAGGCATCAGAGAAGTGAATTGACTCAGGGCATCTGCCTGGAG 666

Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 667 CAAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAAGACAGGGAGGGGCTGGCCCGA 726

Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 727 GTGCTGTGTGGGGAGGACGAGCGTGATGCTGATGCTGGGGCCCAAGGTATGCTCGCTGTC 786

Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
Db 787 CTTGCCACGTCTGAGTGGAGCCCTCAGTGTCTACTGTGGTCTTGGCCCAACAGAACAGAA 846

Qy 241 IleSerSertyLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 847 ATTTCCAGCAAACTCCAACCTTATGAAAAGACCACTCTGACCTGAAAACGCTGGGATC 906

Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 907 CTGATTTTCACTGACGAAGATGTTGCAAGCCACAGAGCTATTATCCCAAAAGACCTGATT 966

Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 967 GCACTGGTCACTCGGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTCTCTGATG 1026

Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
Db 1027 AATGCCGACCTGGAGCCCCACAGGAAAGCTGGCGGAGACCCT 1074

RESULT 4
AAF20898
ID AAF20898 standard; DNA; 2615 BP.
XX
AC AAF20898;
XX
DT 14-MAR-2001 (first entry)
DE Human CD-34 polynucleotide fragment #2465.
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN W0200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (DYEC-) UNIV EAST CAROLINA.
XX
PI (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 292-293; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),

CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.

XX SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:

Pred. No.: 5.28e-115 Length: 2615
Score: 1574.00 Matches: 313
Percent Similarity: 99.37% Conservative: 1
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 98.93% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-4 (1-316) x AAF20898 (1-2615)

Qy 1 MetProArgGlyThrPThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 294 ATGCGCGGGGCTGGACCGGCTTGTGCTGAGTTGCTGCTTGGGTTTCATGAGT 353
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 354 CTTGACACACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAAAATGTT 413
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 414 TCTACAATGTATCTACCAAGAATCTACACACCTAGTACCTTGGAGTACCGAGCTG 473
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 474 CACCCTGTGTTCAACATGGCAATGAGGCCACAAACATCAGAGAACGACAGTCAAA 533
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 534 TTCACATCTACCTGTGATAACCTCAGTTTATGAAACACAAACTCTTCTGTCCAGTCA 593
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 594 CAGACCTCTGTATACACACAGTGTTCACCCACCCACCAACGTTTCAACTCCAGAGACA 653
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 654 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTACAGACCTTTCAACACCTAGCAGTACG 713
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 714 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCCTATCTTAAGTGACATCAAG 773
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 774 GCAGAAATCAATGTTTACGAGCATCAGAGAAGTGAATTTGACTAGGCGATCTGCTGGAG 833
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyLeuAlaArg 200
Db 834 CAAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAGACAGGGGAGGCGCTGCCCGGA 893
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlnValCysSerLeuLeu 220
Db 894 GTGCTGTGTGGGGAGGAGCGAGCTGTGCTGATGCTGGGGCCAGGATGCTCCCTGCTC 953
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 954 CTTGCCAGTCTGAGGTGAGGCGCTCAGTGTCTACTGTGTGCTTGGCCACAGAACAGAA 1013
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 1014 ATTTCCAGCAAACTCCAACTTATGAAAAGACCACTGACCTGAAAGACCTGGGATC 1073
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1074 CTAGATTTCTACTGACGAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAGACCTGATT 1133

Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 1134 GCACCTGTGCTACCTGGGAGCCCTGCTGCTCTTGGCATCATCTGCTATTTCTGATG 1193
Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
Db 1194 AATCGCGGAGCTGGAGCCACAGGAGAAAGGCTGGGCGAGACCT 1241

RESULT 5

AAA34776
ID AAA34776 standard; DNA; 2615 BP.

XX AC AAA34776;

XX DT 28-JUL-2000 (first entry)

XX DE Human adenosine receptor related polynucleotide SEQ ID NO:2465.

XX KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiasthmatic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX OS Homo sapiens.

XX PN WO200009525-A2.

XX PD 24-FEB-2000.

XX PF 03-AUG-1999; 99WO-US17712.

XX PR 03-AUG-1998; 98US-0095212.

XX PA (UYEC-) UNIV EAST CAROLINA.

XX PI Nyce JW;

XX WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
XX vasoconstriction, inflammation, allergies, asthma, hypertension,
XX bronchitis, emphysema, respiratory distress syndrome, ischemia or
XX cancers -

XX Disclosure; Page 602; 1343pp; English.

XX The present invention describes a new composition comprising an
XX antisense oligonucleotide (ON) with low adenosine (up to 1%), which
XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
XX inflammation. The ON can have antiinflammatory, antiallergic,
XX antiasthmatic, cytostatic and analgesic activities. The compositions are
XX useful for the treatment of diseases associated with inflammation,
XX impaired airways, including lung disease and diseases whose secondary
XX effects afflict the lungs of a subject. They can be used for treating
XX e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
XX asthma, impaired respiration, respiratory distress syndrome, pain, cystic
XX fibrosis, pulmonary hypertension, emphysema, chronic obstructive
XX pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
XX carcinomas, and cancers which may metastasise to the lungs, including
XX breast and prostate cancer. The reduction of the adenosine content of
XX the ONs reduces side effects. The A-containing ONs break down with the
XX release of deoxyadenosine which activates adenosine receptors causing
XX bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
XX nucleotide sequences given in the sequence listing from the present
XX invention, which correspond to SEQ ID NO:1 to 2815, and then the last
XX 185 sequences are also called SEQ ID NO:1 to 185, and then the sequences
XX differ from the previously named sequences. SEQ ID NO:11 to 1680
XX (AAA32323 to AAA33992) are specifically claimed ONs from the present
XX invention. N.B. Sequences given in the disclosure of the present

CC invention do not match up with their corresponding SEQ ID NO: sequences
XX given in the sequence listing.

SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:

Pred. No.: 5,28e-115 Length: 2615
Score: 1574.00 Matches: 313
Percent Similarity: 99.37% Conservative: 1
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 98.93% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-4 (1-316) x AAA34776 (1-2615)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
DB 294 ATGCCCGGGGTGGACCGCTTTCCTGCTGAGTTTCTGCTTCTGGGTTCATGAGT 353
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 354 CTTGACAAACAGGTACTGCTACCCAGAGTTACTTACCAGGGACATTTTCAATGTT 413
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 414 TCTACAAATGTATCCTTACCAGAAACTACACACACCTAGTACCCTTGGAAAGTACCAGCCTG 473
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
DB 474 CACCCTGTCTCAACATGCGCAATGAGGCCACACAAACATCACAGAAACGAGTCAAA 533
QY 81 PheThrSerThrSerValIleThrSerValTyrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 534 TTCACATCTACCTCTCTGTATACCTCAGTTTATGGAACACAAACTCTTCTGTGTCAGTCA 593
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 594 CAGACCTCTGTATATCAGCACAGTGTTCACCACCCACGCAACGTTTCAACTCCAGAGACA 653
QY 121 ThrLeuLeuProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
DB 654 ACCTTGAAGCCTAGCCTGTCACTGGAAATGTTTCAACCTTTCACCACTAGCAGCTAGC 713
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
DB 714 CTTGCAACATCTCCCACTAAACCTTATACATCATCTTCTCCTATCTTAAGTGACATCAAG 773
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
DB 774 GCAGAAATCAATGTTCAGGCATCAGAGAAGTGAAATTGACTCAGGGCATCTGCCTGGAG 833
QY 181 GlnAspLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
DB 834 CAATAATAGACCTCCACGCTGCGGAGTTTAAAGAGCAGGGAGAGGCGCTGGCCCCGA 893
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
DB 894 GTGCTGTGTGGGAGGAGCAGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCTC 953
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
DB 954 CTTGCCACGCTGTGAGTGAGGCGCTCAGTGTCTACTGTCTGTCTGTGGCCAAACAGAGAA 1013
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
DB 1014 ATTTCCAGCAATCTCAACTATGAAAGACCAATCTGACCTGAAAGAGCTGGGGATC 1073
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
DB 1074 CTGATTTTCACGTGACAGAGATGTTGCAAGCCACAGAGTATTTCCAAAGACCTGATT 1133
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300

DB 1134 GCACCTGGTCCACCTCGGAGCCCTGCTGGCTGTTTGGGCATCCTACCTGGCTTATTCCTGATG 1193
QY 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
DB 1194 AATCGCCGAGCTGGAGCCCAAGAGAGAGCTGGCGAAGACCCCT 1241
RESULT 6
AAF20899
ID AAF20899 standard; DNA; 3490 BP.
XX AAF20899;
XX AAF20899;
DT 14-MAR-2001 (first entry)
XX Human CD-34 polynucleotide fragment #2466.
XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
cancer; ss.
XX Homo sapiens.
OS Homo sapiens.
XX WO2000062736-A2.
PD 26-OCT-2000.
XX 24-MAR-2000; 2000WO-US08020.
XX 06-APR-1999; 95US-0127958.
PA (UYEC-) UNIV EAST CAROLINA.
XX (NYCE/) NYCE J W.
PI Nyce JW;
XX WPI; 2000-679539/66.
DR Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX Disclosure: Page 291-292; 1592pp; English.
XX The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome

CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
xx
SQ Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;

Alignment Scores:
Pred. No.: 7.68e-115 Length: 3490
Score: 1574.00 Matches: 313
Percent Similarity: 99.37% Conservative: 1
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 98.93% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-4 (1-316) x AAF20899 (1-3490)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1169 ATGCCGGGGCTGACCGCGCTTGTCTGTGAGTTTGGCTTCTGGGTTTCATGAGT 1228
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 1229 CTTGACAACAACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAATGTT 1288
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 1289 TCTCAATGTATCTACCAAGAACTACAACACTAGTACCTTGGAAAGTACCAGCCTG 1348
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 1349 CACCTGTGTCTCAACATGGCAATGAGCCACAAACATCACAGAAACGACAGTCAAA 1408
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 1409 TTCACATCTACCTCTGTGATAACCTCAGTTATGGAACACAAACTCTTCGTCCAGTCA 1468
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 1469 CAGACCTCTGTAATCAGCACAGTGTTCACCACCCAGCAACGTTTCAACTCCAGAGACA 1528
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrSerThrSer 140
Db 1529 ACCTTGAGCCTACGCTGTACCTGGAATGTTTCAGACCTTCAACACCTAGCAGCTAGC 1588
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 1589 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTTCTCTATCTTAAGTGACATCAAG 1648
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 1649 GCAGAAATCAAAATGTTACGAGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCTGGAG 1708
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 1709 CAAATAAGACTCCAGCTGTGCGAGTTTAAGAGGACAGGGAGAGGGCTGCCCCGA 1768
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 1769 GTGCTGTGTGGGAGGACAGCGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCTC 1828
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
Db 1829 CTTGCCAGCTGAGGTGAGCGCTCAGTGTCTACTGTGCTGTTGGCCCAACAGACAGAA 1888
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 1889 ATTTCCAGCAACCTCCACTTATGAAAGACCACTATGACCTGAAAGAGCTGGGGATC 1948
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280

Db 1949 CTAGATTTCCTACTGAGCAAGATGTTTCAAGCCACAGAGCTATTCCTCAAAAGACCTGATT 2008
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 2009 GCATGTGTCACTCGGAGCCCTGCTGGCTCTCTTGGGCATCACTGGCTATTCTCTGATG 2068
Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuLeuGluLeuPro 316
Db 2069 AATGCCGCGACTGGAGCCCTGAGCCCTCACAGAGGAAGGCTGGCGGAGACCT 2116

RESULT 7
AAA34777
ID AAA34777 standard; DNA; 3490 BP.

XX AAA34777;

XX 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide SEQ ID NO:2466.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.

XX WO200009525-A2.

XX 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -

PS Disclosure; Page 603; 1343pp; English.

XX The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680

CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.

XX
SQ Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;

Alignment Scores:
Pred. No.: 7.68e-115 Length: 3490
Score: 1574.00 Matches: 313
Percent Similarity: 99.37% Conservative: 1
Best Local Similarity: 98.05% Mismatches: 2
Query Match: 98.93% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-4 (1-316) x AAA34777 (1-3490)

Qy 1 MetProArgGlyTrpPhrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1169 ATCCCGGGGGCTGGACCGCGCTTCTGCTGAGTTTGTGCTGCTTCTGGGTTTCATGAGT 1228
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 1229 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 1288
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 1289 TCTACAAATGTATCTTACCAAGAACTACAAACCTAGTACCTTGGAGTACCAAGCTG 1348
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 1349 CACCCCTGTGCTCAACATGGCAATGAGGCCACCAACAACATCACAGAAAGACAGTCAA 1408
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 1409 TTTACATCTACCTCTGTGATAACCTCAGTTTATGGAACAACAACCTCTCTGTCCAGTCA 1468
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 1469 CAGACCTCTGTATACAGCAGAGTTTCACCCAGCCAGCAACCTTCACTCCAGAGACA 1528
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 1529 ACCTTGAAGCCTAGCCTGTCCACCTGGAAATGTTTCAGACCTTTCACACCTAGCATTGC 1588
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 1589 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCCTAAGTGACATCAAG 1648
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 1649 GCAGAAATCAAAATGTTCCAGGCATCAGAGAGTCAAAATTGACTCAGGCGCATCTGCCTGGAG 1708
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 1709 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGAGGGGAGGCGCTGGCCCGA 1768
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 1769 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGTGGGCGCCAGGATGCTCCCTGCTC 1828
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
Db 1829 CTTGCCAGTCTCAGGTGAGGCGCTCAGTGTCTACTGCTGTGCTTGTGCCACACAGAGAA 1888
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 1889 ATTTCCAGCAAACTCCAACTTATGAAAGACCACTCTGACCTGAAAGAGCTGGGATC 1948
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1949 CTAGATTTCAGTGAAGAATGTTGCAAGCCACAGAGCTATTTCCTCCAAAGAGCCCTGATT 2008

Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 2009 GCACGTGTCACCTCGGAGCCCTGCTGGCTGTCTGGGCATCATCTTCTCTGATG 2068
Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
Db 2069 AATCGCCGAGCTGGAGCCGCCACAGAGAAAGGCTGGCGGAGACCT 2116

RESULT 8
AAD22663

ID AAD22663 standard; DNA; 906 BP.

XX AAD22663;

XX 26-FEB-2002 (first entry)

XX Human CD34 surface antigen deleted variant (dCD34) encoding DNA.

XX Human; surface marker; surface antigen; T lymphocyte; gene therapy;
KW CD34 deleted variant; dCD34; ds.

XX Homo sapiens.
OS Synthetic.

FH Key Location/Qualifiers
CDS 1..906
FT /*tag= a
FT /product= "Human CD34 surface antigen
FT deleted variant (dCD34)"
XX

PN EP1148066-A1.

XX 24-OCT-2001.

XX 18-APR-2001; 2001EP-0109374.

XX 18-APR-2000; 2000DE-1019075.

XX (ZAND/) ZANDER A R.

XX Zander AR;

XX WPI; 2002-019289/03.

DR P-PSDB; AAE13543.

XX New gene transfer vector (accession number DSM13396) containing a
transgene and a nucleic acid sequence coding for a surface marker,
useful in gene therapy, and for detecting genetically modified cells or
cells which do not express CD34 -

XX Claim 3; Page 14-15; 28pp; English.

XX The patent discloses a gene transfer vector (accession number DSM13396)
containing a transgene and a nucleic acid sequence coding for a surface
marker. The surface marker is the CD34 surface antigen, its fragment or
variant. The vector is useful for in vitro transduction of T lymphocytes,
for gene therapy, and in the enrichment, detection and analysis of cells
in vitro that do not naturally express CD34. T lymphocytes transduced
with the vector are also useful in gene therapy. The CD34 nucleic acid
sequences (marker genes), their fragments or variants are used for
detecting genetically modified cells or cells which do not naturally
express CD34. The present sequence is a DNA encoding human CD34 surface
antigen deleted variant (dCD34).

SQ Sequence 906 BP; 247 A; 250 C; 199 G; 210 T; 0 other;

Alignment Scores:

Pred. No.: 1.84e-110 Length: 906
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 94.85% Indels: 0
DB: 24 Gaps: 0

ID ABA51134 standard; DNA; 233 BP.
XX
AC ABA51134;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #9829.
XX
KW Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 4; SEQ ID NO 9829; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and Bt 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at fcp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. No.: 1.71e-16 Length: 233
Score: 311.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.55% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-4 (1-316) x ABA51134 (1-233)

QY 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
DB 231 TCtGGGTTCATGAGTCTTGACACACACGGTACTGCTACCCAGAGTTACTACCCAGGGA 172

QY 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeu 55
DB 171 ACATTTTCAATGTTCTCAAAATGTATCTACCAAGAACTACAAACACCTAGTACCCCT 112
QY 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
DB 111 GGAAGTACCAGCGCTGCACCTGTGTCTCAACATGGCAATGAGGCCACACAAACATCACA 52
RESULT 12
ABA69132/c
ID ABA69132 standard; DNA; 233 BP.
XX
AC ABA69132;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #17437.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
PS Claim 4; SEQ ID NO 17437; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at fcp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. No.: 1.71e-16 Length: 233
Score: 311.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.55% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-4 (1-316) x ABA69132 (1-233)

QY 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
DB 231 TCtGGGTTCATGAGTCTTGACACACACGGTACTGCTACCCAGAGTTACTACCCAGGGA 172

Qy	36	ThrPheSerAsnValSerThrIleAsnValSerTyrGlnGluThrThrThrProSerThrLeu	55
Db	171	ACATTTTCAAAATGTTTCTCAAAATGTATCTACCAAGAACTACAAACACCTAGTACGCTT	112
Qy	56	GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr	75
Db	111	GGAATGACAGCCCTGCACCCCTGTGTCTCAACATGCAATGAGGCCAACCAACATCACA	52
RESULT 13			
ABA36064/c			
ID	ABA36064	standard; DNA; 233 BP.	
XX	AC	ABA36064;	
AC	AC		
DT	DT		
XX	23-JAN-2002	(first entry)	
XX		Probe #14530 for gene expression analysis in human heart cell sample.	
DE			
XX		Human: gene expression; heart; microarray; vascular system; probe;	
KW		cardiovascular disease; hypertension; cardiac arrhythmia;	
KW		congenital heart disease; ss.	
XX			
XX		Homo sapiens.	
OS			
PN		WO200157274-A2.	
XX			
XX		09-AUG-2001.	
PD			
XX			
XX		30-JAN-2001; 2001WO-US00666.	
XX			
XX		04-FEB-2000; 2000US-0180312.	
PR		26-MAY-2000; 2000US-0207456.	
PR		30-JUN-2000; 2000US-0608408.	
PR		03-AUG-2000; 2000US-0632366.	
PR		21-SEP-2000; 2000US-0234687.	
PR		27-SEP-2000; 2000US-0236359.	
PR		04-OCT-2000; 2000GB-0024263.	
XX			
PA		(MOLE-) MOLECULAR DYNAMICS INC.	
XX			
XX		Penn SG, Hanzel DK, Chen W, Rank DR;	
PI			
XX		WPI; 2001-488899/53.	
DR			
XX		Single exon nucleic acid probes for analyzing gene expression in human	
PT		hearts -	
PT			
XX			
XX		Claim 4; SEQ ID No 14530; 530pp; English.	
PS			
XX			
CC		The present invention relates to single exon nucleic acid probes for	
CC		measuring human gene expression in a sample derived from human heart. The	
CC		present sequence is one such probe. The probes may be used for	
CC		predicting, measuring and displaying gene expression in samples derived	
CC		from the human heart via microarrays. By measuring gene expression, the	
CC		probes are useful for predicting, diagnosing, grading, staging,	
CC		monitoring and prognosing diseases of the human heart and vascular system	
CC		e.g. cardiovascular disease, hypertension, cardiac arrhythmias and	
CC		congenital heart disease.	
CC		Note: The sequence data for this patent did not form part of the printed	
CC		specification, but was obtained in electronic format directly from WIPO	
CC		at ftp.wipo.int/pub/published_pct_sequences .	
XX			
XX		Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;	

US-09-836-602-4 (1-316) x ABA36064 (1-2333)

Qy	16	SerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGly	35
Db	231	TCTGGGTTTCATGAGTCTTGACAAACAGGGTACTGCTACCCAGAGTTACTACCCAGGA	172
Qy	36	ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu	55
Db	171	ACATTTTCAATGTTTCTACAAATGTATCTACCAAGAACTACAACACCTAGTACCCCTT	112
Qy	56	GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr	75
Db	111	GGAAGTACACGCTGCACCTGTGTCTCAACATGGCAATGAGGCCACACAACATCACA	52
RESULT 14			
AAK17442/c			
ID	AAK17442	standard; DNA; 233 BP.	
XX	AC	AAK17442;	
XX	DT	05-NOV-2001 (first entry)	
XX	XX	Human brain expressed single exon probe SEQ ID NO: 17433.	
DE	XX		
XX	XX	Human; brain expressed exon; gene expression analysis; probe;	
KW	KW	microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;	
KW	KW	epilepsy; cancer; ss.	
OS	XX	XX	
OS	XX	Homo sapiens.	
XX	XX		
PN	XX	WO200157275-A2.	
XX	XX	09-AUG-2001.	
PD	XX		
PF	XX	30-JAN-2001; 2001WO-US00667.	
XX	XX		
PR	XX	04-FEB-2000; 2000US-0180312.	
PR	XX	26-MAY-2000; 2000US-0207456.	
PR	XX	30-JUN-2000; 2000US-0608408.	
PR	XX	03-AUG-2000; 2000US-0632366.	
PR	XX	21-SEP-2000; 2000US-0234687.	
PR	XX	27-SEP-2000; 2000US-0236359.	
PR	XX	04-OCT-2000; 2000GB-0024263.	
XX	XX	(MOLE-) MOLECULAR DYNAMICS INC.	
PA	XX		
PI	XX	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX	XX		
XX	XX	WPI; 2001-483446/52.	
XX	XX		
PT	XX	Single exon nucleic acid probes for analyzing gene expression in human	
PT	XX	brains -	
XX	XX		
PS	XX	Example 4; SEQ ID NO: 17433; 650pp + Sequence Listing; English.	
XX	XX		
CC	XX	The present invention provides a number of single exon nucleic acid	
CC	XX	probes which are derived from genomic sequences expressed in the human	
CC	XX	brain. They can be used to measure gene expression in brain cell samples	
CC	XX	which may enable the diagnosis and improved treatment of nervous system	
CC	XX	diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,	
CC	XX	epilepsy and cancers. The present sequence is one of the probes of the	
CC	XX	invention.	
XX	XX		
SQ	XX	Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;	
Alignment Scores:			
Pred. No.:		1-71e-16	Length: 233
Score:		311.00	Matches: 60
Percent Similarity:		100.00%	Conservative: 0
Best Local Similarity:		100.00%	Mismatches: 0
Query Match:		19.55%	Indels: 0
DB:		22	Gaps: 0
DB:			

US-09-836-602-4 (1-316) x AAK17442 (1-233)

QY	16	SerClypPheMetSerLeuAspAsnGlyThrAlaThrProCLeuPonThrGlnGly	35
DB	231	TCYGGGTTCATGAGTCTTGACAAACAGGTACTGCCACCCAGAGTTACTACCCAGGA	172
QY	36	ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu	55
DB	171	ACATTTTCAAAATGTTCTCAAAATGTATCCTACCAAGAACTACAAACCTAGTACCCCT	112
QY	56	GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr	75
DB	111	GGAAGTACCAAGCTGCACCTGTGTCTCAACATGGCAATGAGGCCACAACAACATCACA	52

RESULT 15

RESOL 13
AAK43242/C

ID AAK43242 standard: DNA: 233 BP.

AA
AC AAK43242:XX
DT 06-NOV-20

XX

XX

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX
===== / Infancia / infancia / infancia /

XX
XX

XX

PN WO200157276-A2.
YY

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 30-JUN-2000; 2000DS-0608408.

PK 03-AUG-2000; 2000US-0632388;
PR 21-SEP-2000; 2000US-0234687;

PR 27-SEP-2000; 2000S-0236359.
BR 04-OCT-2000; 2000CB-0031363

XX

XX

XX
XXXX
XX

— 3 —

PT analyzing gene expression in human bone

XX
PC
Example 1: CEO TD NO

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

CC probes which are derived from genomic sequences expressed in the human

CC samples, which may enable the improved diagnosis and treatment

CC such as lymphoma, leukaemia and myeloma. The present sequence is one of the probes of the invention.

Sequence 233 BP: 55 A: 47 C: 61 G: 70 T: 0 other:

Alignment Scores:

Pred. No.:	1.71e-16	Length:	233
Count:	217	Count:	2

Percent Similarity: 100.00% Conservative: 0

Query Match:	19.55%	Indels:	0
--------------	--------	---------	---

3
2
1

(cc2-T) 247264VH Y (OTC-T) 4-200-058 60 60

[illegible]

Search completed: October 30, 2002, 08:25:46

Job time : 152.956 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:08:08 ; Search time 1619.9 Seconds
(without alignments)
4082.221 Million cell updates/sec

Title: US-09-836-602-4
Perfect score: 1591
Sequence: 1 MPRGWTALCLLSLLPSGFMS.....YFLMNRKSWSPGTGERLELEP 316

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first 45 summaries

Post-processing: Minimum Match 0%

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09836602/runat_29102002_101144_3785/app-query.fasta_1.1429
-DB=GenEmbl -QMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602.ecgn_1.1.3800 -runat_29102002_101144_3785 -NCPU=6 -ICPU=3
-NO_XLUPY -NO_XMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.pl.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.or.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1591	100.0	2657	6	AX333753	AX333753 Sequence
2	1591	100.0	2657	9	S53911	S53911 CD34-glycop
3	1574	98.9	2463	9	S53910	S53910 CD34-glycop
4	1574	98.9	2615	9	HUMCD34HS	M81104 Human CD34
5	964	60.6	2956	4	CFU49457	U49457 Canis faml
6	921	57.9	1065	6	AR078454	AR078454 Sequence
7	915	57.5	2690	4	AF461503	AF461503 Sus scrofa
8	887.5	55.8	978	10	S69301	S69301 CD34+cell s
9	870.5	54.7	1149	10	S69299	S69299 CD34+cell s
10	870.5	54.7	1260	10	S69293	S69293 CD34+stem c
11	870.5	54.7	2398	10	BC006607	BC006607 Mus muscu
12	793	49.8	581	9	AF202879	AF202879 Homo sapi
13	772.5	48.6	2453	4	AB021662	AB021662 Bos tauru
14	741.5	46.6	2420	4	AF259378	AF259378 Capra hir
15	617	38.8	160771	9	HS88L2	AL035091 Human DNA
16	617	38.8	182742	2	AC084393	AC084393 Homo sapi
17	617	38.8	185162	2	AL365178	AL365178 Homo sapi
18	617	38.8	212535	2	AL356275	AL356275 Homo sapi
19	409	25.7	254	9	HUMCD34S3	M81940 Human CD34
20	374.5	23.5	98347	2	AL513203	AL513203 Mus muscu
21	374.5	23.5	166333	2	AL513470	AL513470 Mus muscu
22	311	19.5	183	9	HUMCD34S2	M81939 Human CD34
23	275	17.3	165	9	HUMCD34S7	M81942 Human CD34
24	260	16.3	157	9	HUMCD34S5	M81944 Human CD34
25	233	14.6	235	6	AX247364	AX247364 Sequence
26	172.5	10.8	3124	5	MMTHRAD1	Y13978 Gallus gall
27	167	10.5	80272	5	AL591180	AL591180 Zebrafish
28	166	10.4	89232	5	AL591175	AL591175 Zebrafish
29	162.5	10.2	55069	8	Y5CH9196	U11583 Saccharomyc
30	162.5	10.2	184427	14	EHVU20824	U20824 Equine herp
31	160.5	10.1	253273	2	AL669823	AL669823 Mus muscu
32	158.5	10.0	183358	9	AC019041	AC019041 Homo sapi
33	155.5	9.8	2161	5	XELFIMC1X	L02115 Frog integu
34	155.5	9.8	167390	9	AC007263	AC007263 Homo sapi
35	155.5	9.8	182534	10	AC074046	AC074046 Mus muscu
36	154	9.7	3053	8	SCYNR044W	Z71659 S.cerevisia
37	154	9.7	3137	8	YSCAAGLCS	M60590 S.cerevisia
38	153.5	9.6	167254	9	CNS057DS	AL357093 Human chr
39	153.5	9.6	172789	2	AC104020	AC104020 Homo sapi
40	153	9.6	165227	2	AC098014	AC098014 Rattus no
41	153	9.6	185337	2	AC095762	AC095762 Rattus no
42	153	9.6	286060	2	AL589862	AL589862 Homo sapi
43	152.5	9.6	196553	2	AC073946	AC073946 Mus muscu
44	152	9.6	12412	3	CELH43E16	AC006669 Caenorhab
45	152	9.6	131274	2	AC096869	AC096869 Rattus no

ALIGNMENTS

RESULT 1
AX333753
LOCUS AX333753
DEFINITION Sequence 4262 from Patent WO0194629.
ACCESSION AX333753
VERSION AX333753.1 GI:18124472
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

linear PAT 09-JAN-2002
2657 bp DNA

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (sites)
Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets

JOURNAL

Patent: WO 0194629-A 4262 13-DEC-2001;
Avalon Pharmaceuticals (US)

FEATURES

source
Location/Qualifiers
1..2657
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 624 a 790 c 609 g 634 t
ORIGIN

Alignment Scores:

Pred. No.: 2.58e-97 Length: 2657
Score: 1591.00 Matches: 316
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-836-602-4 (1-316) x AX333753 (1-2657)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
DB 127 ATGCCGGGGGCTGGACCGGCTTGTCTGCTGAGTTGCTTGGCTTGCATAGT 186
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 187 CTTGACAAACAGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTCAAATGTT 246
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 247 TCACAAATATCTCTACCAAGAACTACACACCTAGTACCTTGGAGTACGAGCTG 306
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
DB 307 CACCCTGTGCTCAACATGCGAATGAGGCCACAAACATCACAGAAACGACAGTCAA 366
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 367 TTCACATCTACCTGTGTGATAAGCTCAGTTTATGGAAACACAACTTCTGTCCAGTCA 426
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 427 CAGACCTCTGTATACGACAGTGTTCACCCACCCAGCAACGTTTCAACTCCAGAGACA 486
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
DB 487 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTCAGACCTTTCACACCTAGCCTAGC 546
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
DB 547 CTTGCAACATCTCCCACTAAACCCCTATACATCTCTCTCTATCTTCAAGTCAATCAAG 606
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
DB 607 GCAGAAATCAATGTTTCAGGCATCAGAGAAGTGAATGACTCAGGCGATCTGCCTGGAG 666
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
DB 667 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGGAGGAGGCGCTGGCCGA 726
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
DB 727 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGCCCGGATGCTCCCTGCTC 786
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
DB 787 CTTGCCCGAGTCTGAGGTGAGGCGCTCAGTGTCTACTGCTGTGCTTTGGCCAAACAGAA 846

QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
DB 847 ATTTCAGCAAACTCCCAACTTATGAAAAAGACCAATCTGACCTGAAAAAGCTGGGATC 906
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
DB 907 CTAGATTTCACTGAGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAGACCTGATT 966
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
DB 967 GCATCTGGTCACTCGGAGCCCTGCTGCTTGGGCATCATCTGGCTATTCTCTCATG 1026
QY 301 AsnArgArgSerTyrPheThrGlyGluArgLeuGluLeuGluPro 316
DB 1027 AATCGCCGACGCTGGAGCCCAACAGAGAAAGGCTGGAGCTGGACCC 1074
RESULT 2
LOCUS S53911 2657 bp mRNA linear PRI 08-MAY-1993
DEFINITION CD34=glycoprotein expressed in lymphohematopoietic progenitor cells
[alternatively spliced, truncated form] (human, UT7, mRNA, 2657
nt).
ACCESSION S53911
VERSION S53911.1 GI:264768
KEYWORDS human UT7.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2657)
AUTHORS Nakamura, Y., Komano, H. and Nakauchi, H.
TITLE Two alternative forms of cDNA encoding CD34
JOURNAL Exp. Hematol. 21 (2), 236-242 (1993)
MEDLINE 93146100
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI gibbsq 124009] from the original journal article.
This sequence comes from Fig. 1AB.
FEATURES
Location/Qualifiers
1..2657
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2657
/gene="CD34"
91..1077
/gene="CD34"
/note="glycoprotein expressed in lymphohematopoietic
progenitor cells; This sequence comes from Fig. 1AB"
/codon_start=1
/product="CD34"
/protein_id="AAB25223.1"
/db_xref="GI:264769"
/translation="MLVRRGARAGPMPRGWTALCLLSLPSEMSLDNNGTATPELP
TQGTSPNVSTNVSYQETTPSLGSLHPVSOHNGNEATNITETVKTSTSVTSY
YNTSNSSVQSQTSVISTVETTPANVSTPETTLKPSLSPGNVSDLSSTLSLATSPKP
YTSPPSLDIKAEIKCGIRVKLTQGLKQNTSSCAEFKDRGEGLARVLGCEE
QADAGAQVCSLLLAQSEVPQCLLVLANRTEISSKLQMLMKHQSDLKLGILDT
EDVASHQSQYSQKTLIALVTSGALLAVLGIYFLMNRSSWSPTGERLEUP"
BASE COUNT 624 a 790 c 609 g 634 t
ORIGIN
Alignment Scores:
Pred. No.: 2.58e-97 Length: 2657
Score: 1591.00 Matches: 316
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-836-602-4 (1-316) x S53911 (1-2657)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
DB 127 ATGCCGGGGGCTGGACCGGCTTGTCTGCTGAGTTGCTTGGCTTGCATAGT 186

127	ATGCCGGGGCGCTGGACCGCGCTTTCGCTGCTGAGTTTGCCTCTTCGGGTTCATGAGT	186
21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
187	CTTGACAAACAGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTCAAATGTT	246
41	SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu	60
247	TCTACAAATGTATCTACCAAGAACTACAAACACTAGTACCTTGGAAAGTACACAGCTG	306
61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValIys	80
307	CACCTGTGCTCAACATGCAATGAGCCACACAACAACATCACAGAAACACAGTCAAA	366
81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
367	TTACATCTACCTGTGTGATAACCTCAGTTTATGGAACAACAACTCTCTGTCCAGTCA	426
101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
427	CAGACCTCTCTAATACACAGAGTTTCACCAACCCACCAACAGTTTCACTCCAGACAC	486
121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
487	ACCTTGAAGCCTACCTCTCACCTGGAAATGTTTCAGACCTTTCAACCACTAGCACATGC	546
141	LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
547	CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCCTTAGTGACATCAAG	606
161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
607	GCAGAAATCAAAATGTTCAGGCATCAGAGAAAGTGAATGACTCAGGGCATCTGCCTGGAG	666
181	GlnAsnLysThrSerSerCysAlaGluPheLysLysLysAspArgGlyGluGlyLeuAlaArg	200
667	CAAAATGAAGCCTCCAGCTGTGCGAGCTTAAAGAGACAGGGGAGAGGCGCTGGCCCGA	726
201	ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
727	GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGCCCCAGATGCTCCTGCTC	786
221	LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
787	CTTGCCCAAGTCTGAGGTGAGGCCCTCAGTGTCTACTGTGCTTGTGGCCACAGAAACAGAA	846
241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle	260
847	ATTTTCCAGCAAACTCCAACTTATGAAAGACACCAATCTGACCTGAAAGAGCTGGGATC	906
261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle	280
907	CTAGATTTCAGTGAGCAAGATGTGCAAGCCACAGAGCTATTCCCAAAAGACCTGATT	966
281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet	300
967	GCATGTGTCACCTCGGAGGCCCTGCTGGCTGCTCTGGGCATCACTGGCTATTTCCTGATG	1026
301	AsnArgArgSerTrpSerProThrGlyGluArgLeuLeuGluPro	316
1027	AATGCCGAGCTGGAGCCCAACAGAGAAAGGCTGGAGCTGGAAACCC	1074
RESULT 3		
S53910		
LOCUS	S53910	2463 bp mRNA linear PRI 08-MAY-1993
DEFINITION	Cu34-glycoprotein expressed in lymphohematopoietic progenitor cells [alternatively spliced] [human, UT7 cells, mRNA, 2463 nt].	
ACCESSION	S53910	
VERSION	S53910.1	GI:264766
KEYWORDS	human UT7 cells.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 2463)
 Nakamura,Y., Komano,H. and Nakauchi,H.
 Two alternative forms of cDNA encoding CD34
 Exp. Hematol. 21 (2), 236-242 (1993)
 93146100
 GenBank staff at the National Library of Medicine created this
 entry [NCBI bibsq 124007] from the original journal article.
 This sequence comes from Fig. 1A.

FEATURES
 source
 1. .2463
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 gene
 1. .2463
 /gene="CD34"
 CDS
 91..1248
 /gene="CD34"
 /note="glycoprotein expressed in lymphohematopoietic
 progenitor cells; This sequence comes from Fig. 1A"
 /codon_start=1
 /product="CD34"
 /protein_id="AAB25222.1"
 /db_xref="GI:264767"
 /translation="MLVRRGARAGPRMGWATLCLLSLLPSGFMSLDNNGTATPELP
 TQGTFSNVSTVSGTSTPTSTIGSLHPSVSGHGEATNTTETTVKFTSTSVITSP
 YGNTSSVQSVITSVFTVPANVSTPETTLPSLSPGNVSDLTSTTSLATSPKPP
 YTSSTPLSDIKAEIKCSGTFREVKLTOGICLEONKTSSCAEFKKDRGELARVLGGE
 QADADAGAQCSSLLAGSEVYPQCLLVLANRTEISSKLQMLKKHQSDLAKGLGIDPT
 EOVASHQSVSQTLTALVTSGLLAVLGITGYFLNNRWSVPTGERLGEDPYTENG
 GGQVSGPSTSPDAQKASVNGAGTGQATSRNGHSARQHVADTEL"

BASE COUNT 589 a 709 c 580 g 585 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3,19e-96 Length: 2463
 Score: 1574.00 Matches: 313
 Percent Similarity: 99.37% Conservative: 1
 Best Local Similarity: 99.05% Mismatches: 2
 Query Match: 98.93% Indels: 0
 DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x S53910 (1-2463)

Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 |||||
 Db 127 ATGCCGGGGCTGGACCGCCTTGTCTGTGAGTTGCTGCCTTCCTGGGTTCATGAGT 186
 Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 |||||
 Db 187 CTTGACAAACACGGTACTGCTTACCCACAGTACTACCCAGGGAACATTTTCAAAATGTT 246
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 |||||
 Db 247 TCTCAAAATGATCCTACCAAGAAACATACACACCTAGTACCTTGGAAAGTACCAGCCTG 306
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
 |||||
 Db 307 CACCCTGTGCTCAACATGGCAATGAGCCACACAAACATCACAGAACACAGTCAAA 366
 Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 |||||
 Db 367 TTCACATCTACCTCTGTGATACCTCAGTTATTTGGAAACACAAAACCTCTCTGTCCAGTCA 426
 Qy 101 GlnThrSerValIleSerThrValPheThrThrThrProAlaAsnValSerThrProGluThr 120
 |||||
 Db 427 CAGACCTCTGTAATCAGACAGTGTTCACCAACCCAGCAACCTTTCAACTCCAGAGACA 486
 Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 |||||
 Db 487 ACCTTGAAGCCTACGCTGTACCTGGAAATGTTTCAGACCTTTTCACCACTAGCACTAGC 546
 Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
 |||||

Db	547	CTTGCAACATCTCCCACTAAACCCCTATACATCATCTTCTCTCTATCCCTAAAGTGACATCAAG	606
Qy	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	607	GCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATTGACTCAGGGCATCTGCCTGGAG	666
Qy	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	667	CAAAATAGACCTCCAGCTGTCGGAGTTTAAAGAGACAGGGAGAGGGCTTGGCCCGA	726
Qy	201	ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	727	GTGCTGTGTGGGAGGAGCAGGCTGATGCTGCTGGGCCAGGATGCTCCTCTGCTC	786
Qy	221	LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu	240
Db	787	CTTGCCAGTCTGAGGTGAGGCCCTCAGTGTCTACTGTGCTTGGCCACAGACAGAA	846
Qy	241	IleSerSerLysLeuGlnLeuMetLysHisGlnSerAspLeuLysLysLeuGlyIle	260
Db	847	ATTTCCAGCAAACTCCAACCTTATGAAAAGCACCAATCTGACCTGAAAAAGCTGGGATC	906
Qy	261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle	280
Db	907	CTAGATTTCACTGAGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAGACCCCTGATT	966
Qy	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet	300
Db	967	GCATGTGTACCTCGGAGCCCTGCTGGCTGCTCTTGGGCATCATCTGGCTATTTCCTGATG	1026
Qy	301	AsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro	316
Db	1027	AATCGCGCAGCTGGAGCCCCACAGGAGAAAGGCTGGCGGAAGACCCCT	1074
RESULT 4			
LOCUS	HUMCD34HS	2615 bp	mRNA linear PRI 02-NOV-1993
DEFINITION	Human CD34 mRNA, complete cds.		
ACCESSION	M81104.X60172		
VERSION	M81104.1 GI:180108		
KEYWORDS	CD34; hematopoietic stem cell surface antigen; sialomucin.		
SOURCE	Homo sapiens cDNA to mRNA.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2615)		
TITLE	Simmons,D.L., Satterthwaite,A.B., Tenen,D.G. and Seed,B.		
JOURNAL	Molecular cloning of a cDNA encoding CD34, a sialomucin of human		
MEDLINE	hematopoietic stem cells		
FEATURES	J. Immunol. 148, 267-271 (1992)		
source	Location/Qualifiers		
	1..2615		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	294..1415		
	/codon_start=1		
	/product="CD34"		
	/protein_id="AAA03181.1"		
	/db_xref="GI:180109"		
	/translation="MPRGWTAICLLSLPSGFMSLDNNGTATPELPTGCTFFSNVSTNV		
	SVQETPTSLGTSLSHPVSOHQNEATNITETVKFTSTVTSVYGNSTNSVQSQT		
	SVISTVTPANVSPTETLKPSLSPGNVSLSTSTSLATSPTRKPYSSSTPLSDIK		
	AETKSGIREVLKTCICLEQNKTSACAEFKKDRGLARVLCGEGEADADAGAOVCS		
	LLLAQSEVPQCLLVLANETISKKLQKMKHQLDLKLGLIDFTEQDYASHQSVSQ		
	KTLIALVTSGALLVLGITYGYFLMNRKSWSPFGERLGEDPYITENGGGGYSSGGPETS		
	PEAQGKASVNRQAQNGTGGATSRNGHSARQHVADTEL"		
BASE COUNT	616 a	763 c	627 g 609 t
ORIGIN			
Alignment Scores:	3.44e-96	Length:	2615
Pred. No.:	1574.00	Matches:	313
Score:			

Percent Similarity: 99.37%		Conservative: 1	
Best Local Similarity: 99.05%		Mismatches: 2	
Query Match: 98.93%		Indels: 0	
DB: 9		Gaps: 0	
US-09-836-602-4 (1-316) x HUMCD34HS (1-2615)			
Qy	1	MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer	20
Db	294	ATGCCGGGGGCTGGACCGCGCTTTGCTTGTGAGTTTGTGCTTCTGGGTTTCATCAGT	353
Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	354	CTTGACAAACACGGTACTGCTACCCCAAGATTACTACCCAGGGAACATTTTCAAAATGTT	413
Qy	41	SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu	60
Db	414	TCTCAAAATGATCTCTACCAAGAAACTACAAACCTAGTACCTTGGAGTACCAACCTG	473
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
Db	474	CACCTGTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA	533
Qy	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
Db	534	TTCCATCTACTCTGTGATAACCTTCAGTTTATGGAAACACAAACTCTTCTGTCTCCAGTCA	593
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	594	CAGACCTCTGTAATCAGCAGAGTGTCCACCCGCCCAACGTTTCAACTCCAGAGACA	653
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	654	ACCTTGAAGCCTAGCCTGTCACTGGAAATGTTTCAGACCTTTCAACCTAGCACTAGC	713
Qy	141	LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
Db	714	CTTGCAACATCTCCCACTAAACCCATATACATCATCTTCTCTATCTCTTAAGTGACATCAAG	773
Qy	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	774	GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATGACTCAGGGCATCTGCCTGGAG	833
Qy	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	834	CAAAATAGACCTCCAGCTGTGCGAGTTTAAAGAGACAGGAGGAGGCGCTGGCCCGA	893
Qy	201	ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	894	GTGCTGTGTGGGAGGAGCAGGCTGATGCTGCTGCTGGGCCCCAGGATGCTCCTCTGCTC	953
Qy	221	LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
Db	954	CTTGCCCGAGTCTGAGGTGAGGCCCTCAGTGTCTACTGCTGTCTTGGCCCAACAGAACAGAA	1013
Qy	241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle	260
Db	1014	ATTTCCAGCAAACTCCAACTTATGAAAAGCACCAATCTGACCTGAAAAGAGCTGGGGATC	1073
Qy	261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle	280
Db	1074	CTAGATTTCACTGAGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAGACCTGATT	1133
Qy	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValIleGlyIleThrGlyTyrPheLeuMet	300
Db	1134	GCACCTGTGCTACCTCGGAGGCCCTGCTGGCTGTCTTGGGCATCTAGCTATTCTCTGATG	1193
Qy	301	AsnArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro	316
Db	1194	AATCGCCGACCTGGAGCCCCACAGGAGAAAGGCTGGCGGAAGACCT	1241
RESULT 5			
CFU49457			

LOCUS	CFU49457	2956 bp	mRNA	linear	MAM 21-JAN-1997
DEFINITION	Canis familiaris hematopoietic progenitor cell marker CD34 mRNA, complete cds.				
ACCESSION	U09457				
KEYWORDS	U09457.1	GI:1224105			
SOURCE	dog.				
ORGANISM	Canis familiaris				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.				
AUTHORS	McSweeney,P.A., Rouleau,K.A., Storb,R., Bolles,L., Wallace,P.M., Beauchamp,M., Kriзанac-Bengez,L., Moore,P.P., Sale,G., Sandmaier,B., de Revel,T., Appelbaum,F.R. and Nash,R.A.				
TITLE	Canine CD34: cloning of the cDNA and evaluation of an antiserum to recombinant protein				
JOURNAL	Blood	88	(6),	1992-2003	(1996)
MEDLINE	96420219				
REFERENCE	2 (bases 1 to 2956)				
AUTHORS	McSweeney,P.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-FEB-1996) P.A. McSweeney, Transplant Biology, FHCR,				
FEATURES	1124 Columbia St, Seattle, WA 98104, USA				
source	Location/Qualifiers				
	1..2956				
	/organism="Canis familiaris"				
	/db_xref="taxon:9615"				
gene	/tissue_type="ML2 myelomonocytic leukemia"				
	1..2956				
CDS	/gene="CD34"				
	260..1429				
	/gene="CD34"				
	/note="type 1 transmembrane protein"				
	/codon_start=1				
	/product="hematopoietic progenitor cell marker CD34"				
	/protein_id="AAB41055.1"				
	/db_xref="GI:1224106"				
	/translation="MLAGRARAGGGLPRGWALCLSLLPFGTNTVETVTPTTTSTHSAVSNTSKRAITLTSGTTLTVSVSDSGGTATISRTIVHTVTSSEILTTPGMNSQSDTSLATIPNFTNSTSYTLEPPLPGNGSDPPYNSTSLVTSPTETQYTSLSPTSPNDPSPIKIGIKSGVKELNQICILENETSCDDFKKDNEKLQVLCEKPAAEGAGVCILLIAQSVBRPHCLLVLANKEITFSLQLLRKHQSIDLKLGIRDFTEODVGSHQSRKTLIALVTSGLLAVLCTGYFLMNRBSWPTGERLGEDPYTENGGGGYSSGGVSPGAQCKASVNRGPENETGQTATSRNGHSARQHNVADTEL"				
BASE COUNT	625 a	864 c	815 g	652 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	2,17e-55	Length:	2956		
Score:	964.00	Matches:	208		
Percent Similarity:	72.98%	Conservative:	27		
Best Local Similarity:	64.60%	Mismatches:	79		
Query Match:	60.59%	Indels:	8		
DB:	4	Gaps:	2		
US-09-836-602-4 (1-316) x CFU49457 (1-2956)					
Qy	1	MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20		
Dd	296	CTGCCGGGGCGCTGGACCGCCTCTGCCTGCTCACTGCTGCCTTTGGGTTCCAAAAC	355		
Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40		
Dd	356	ACAGAACCCTGATTAATTCTTACCACAGTCGCCAACCTCCACAGAAATAATCTCAGCTGTT	415		
Qy	41	SerThrAsnValSerTyrgIngluThrThrThrProSerThrLeuGlySerThrSerLeu	60		
Dd	416	TCTGAGATACATCCAAACGGGAACCATCACACTAATCTCTTGGAATACCACTG	475		
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVallys	80		
Dd	476	TACTCTGCTCTCAAGACAGCAGTGGGACCAACGACCACTCTCAGACACTACAGTCCAT	535		

Qy	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
Db	536	GTCAATCTTACCTTCTGAGATCACCCCTAACGCGCTGGACCATGAACCTTCTGTGTCACTGC	595
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	596	CAGACCTCTTTAGTCATACAGGTATCTTTTACCCCAACCACTTTTCAACTTCAAGTGTG	655
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	656	ACCTTGGAGCCACGCTGCTACCTGGAATGGTTCCGATCCCCCTACACAGACCCAGC	715
Qy	141	LeuAlaThrSerProThrLysProTyrThrSerSerSerPro	154
Db	716	CTTGTGATCCCCCAGGAATATTATACATCACTTCTCTTACCCCAAGTAGAAATGAC	775
Qy	155	IleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThr	174
Db	776	ACCCCAAGTACCATCAAGGGAGAAATCAATGTTCGGAGTCAAGAAGTGAATTAAC	835
Qy	175	GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArg	194
Db	836	CAGGTATCTGCCTAGAGCTAAATGAGACCTCCAGCTGTGAGGACTTTAGAAAGATAAC	895
Qy	195	GlyGluGlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaGlyAla	214
Db	896	GAAGAAAACTGACCCAAAGTCTCTGT-----GAGAAGGAGCCAGCTGAGGCTGGGCCC	949
Qy	215	GlnValCysSerLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuVal	234
Db	950	GGGTGTGTCTCCCTGCTCTGCGCCAGTCTGAGTGAGGCTCACTGCCCTGCTGCTGTC	1009
Qy	235	LeuAlaAsnArgThrGluIleLeuSerLysLysLeuGlnLeuMetLysLysHisGlnSerAsp	254
Db	1010	TTGCCCAACAAAAACAGAACTTTTCAGTAACTCCAACTTCGAGAAAGACCAGCTCTGAC	1069
Qy	255	LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyr	274
Db	1070	CTGAAAAAGCTGGGATCCGAGACTTCACCTGAACAAGATGTTGGGAGCCACAGAGCTAT	1129
Qy	275	SerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIle	294
Db	1130	TCCCGCAAGACCCCTGATGTCAGTCCGTCACTCAGGGATCCTGCTGGCTGTCTTGGGCACC	1189
Qy	295	ThrGlyTyrPheLeuMetAsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeu	314
Db	1190	ACTGGTACTCTCTGATGAACCGCGGACGATGGAGCCCTACAGAGAGAAAGGCTGGGGCAA	1249
Qy	315	GluPro 316	
Db	1250	GACCT 1255	
RESULT 6			
AR078454			
LOCUS	AR078454	1065 bp	DNA linear PAT 31-AUG-2000
DEFINITION	Sequence 18 from patent US 5962644.		
ACCESSION	AR078454		
VERSION	AR078454.1	GI:10005200	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1065)		
AUTHORS	Hawley, R.J. and Monroy, R.L.		
TITLE	Porcine CD34		
JOURNAL	Patent: US 5962644-A 18 05-OCT-1999;		
FEATURES	Location/Qualifiers		
source	1..1065		
	/organism="unknown"		
BASE COUNT	248 a 326 c 273 g 218 t		
ORIGIN			
Alignment Scores:			

Qy	300	MetAsnArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro	316
Db	1012	ATGAACCGTCGAGTTGGAGCCCTACAGGAGAAAGGCTGGAGCTGGAACCC	1062
RESULT 7			
AF461503		AF461503	
LOCUS		2690 bp	linear
DEFINITION		CD34 antigen mRNA, complete cds.	
ACCESSION		AF461503	
VERSION		AF461503.1	
KEYWORDS		GI:18308137	
SOURCE		pig.	
ORGANISM		Sus scrofa	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Euthera; Cetartiodactyla; Suina; Suidae; Sus.	
TITLE		1 (bases 1 to 2690)	
JOURNAL		Sun, J. and Butler, J.E.	
REFERENCE		2 (bases 1 to 2690)	
AUTHORS		Sun, J., Wang, J. and Butler, J.E.	
TITLE		Direct Submission	
JOURNAL		Submitted (20-DEC-2001) Microbiology, University of Iowa, 51 Newton	
FEATURES		Location/Qualifiers	
source		1..2690	
		/organism="Sus scrofa"	
		/db_xref="taxon:9823"	
		/tissue_type="mesenteric lymph nodes"	
CDS		298..1467	
		/note="cluster of differentiation antigen 34"	
		/codon_start=1	
		/product="CD34 antigen"	
		/protein_id="AAL67838.1"	
		/db_xref="GI:18308138"	
		/translation="MPRGWTTICLLSPGFTAVNSVAVSTPLPTAAGSTPTGTPATA	
		GTATGATSATDISPVSYTNISNETTPDAFESASLHVSGSGSTTVASLGSTVNFMS	
		TSATLVPEVNSVQPSQSLAVSSATINFTTSEVTLPQSTFPFGNVSDPLYNSTPSA	
		RSPTSPVTSPPTPGSHKGVKCAQIKVKLTQIGCLERNETSGCEKFKDKNGEKLMO	
		ILGGEQAEAGPGVCULLLAQSVKPHCLLLVANGTELSSKFLLEKHQSLEKREMSI	
		QHSQDVRHQSHYSKRTLLAVLTGSLIYALVIGITVGLLMNRNRSWSPGRLGEDPY	
		TENGSGQGVSSGPGASVTRQAQNGTGTATSRNGHSARQPMVADTEL"	
BASE COUNT	547 a	808 c	738 g
ORIGIN		597 t	
Alignment Scores:			
Pred. No.:	3.6e-52	Length:	2690
Score:	915.00	Matches:	199
Percent Similarity:	69.44%	Conservative:	35
Best Local Similarity:	59.05%	Mismatches:	77
Query Match:	57.51%	Indels:	26
DB:	4	Gaps:	6
US-09-836-602-4 (1-316) x AF461503 (1-2690)			
Qy	1	MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	298	ATGCCGGGGGCTGGACACAGCTCTGCTGCTGAGTTTGCTGCCCTCTGGGTTTCACAGCT	357
Qy	21	LeuAspAsnAnglyThrAlaThrProGluLeuPro-----	32
		::: :::	
Db	358	GTG---AACAGCTCGACTGTGTCTCCACCTTGCCAACTGCCGCTGGGTCAACTCCACC	414
Qy	33	-----ThrGlnGlyThr-----	39
Db	415	GGGCAGCTACTGCAGGAGACAGCTATCACCGGGTCAGCTATCTCAGACATATCTCACCT	474
Qy	40	ValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSer	59
Db	475	GTTTCTACAATAATATATCCACAGGAGAA---ACCACACCATGCTTTCGAAAGTGCACG	531
Qy	60	LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVal	79

```
Db 532 CTCACACTGTCTCTCAGGGCAGCAGTGGGACCGTAGCCATCTCAGGCCCTCAGATT 591
Qy 80 LysPheThrSerThrSerValIleThrSerValThrGlyAsnThrAsnSerSerValGln 99
Db 592 AATTTCATGTCTACCTCGCGGTACCTCGTCCCGAAGAACTGTTCTCTCCAG 651
Qy 100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
Db 652 CCTGAGACTCT--CTAGCCACAGTGTCTCCGCCACCATCACTTTACAACTTCAGAG 708
Qy 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 139
Db 709 GTGACCTGTCAGCCCGCAGCAGCTCCAGGAATGTTTTCAGACCCCTCTACAACTGACC 768
Qy 140 SerLeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIle 159
Db 769 AGCCCTGCGAGATCCCCACCGCCCTACACATCATCTCTCTTACCCAGGTAGCCAC 828
Qy 160 LysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Db 829 AAGGGGAGTCAATGTGCCAATCAAGAGGTGAATTGACCCAGGTATCTGCCTG 888
Qy 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Db 889 GAGCGAATGAGACCTCCGGCTGCGAGAAGTTTAAAGAGGACAATGGAGAGAAGTTGATG 948
Qy 200 ArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
Db 949 CAATTCCTGTGTGGCAGGAGCAGCTGAGCGCGGCCAGGG-----GTGTGCTCCTGTG 1002
Qy 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThr 239
Db 1003 CTCCTTGCCCAATCTGAGTGAACCTCAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1062
Qy 240 GluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGly 259
Db 1063 GAATCTAGCAGCAAGTCTGCTGCTGGAAGAACCACTGCTGAACTGAGAGAGATGAGC 1122
Qy 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerThrSerGlnLysThrLeu 279
Db 1123 ATCCAACTCTTCGAAACAGATGTTAGGAGCCACCACTGCTGAACTGAGAGAGATGAGC 1182
Qy 280 IleAlaLeuValThrSerGlyAlaLeuAlaValLeuGlyIleThrGlyThrPheLeu 299
Db 1183 ATTGACCTGTCACCTCGGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1242
Qy 300 MetAsnArgSerThrSerProThrGlyGluArgLeuGluLeuGluPro 316
Db 1243 ATGAACCGTCGAGTTGGAGCCCTACAGGAGAAAGCTGGCGAAGACCT 1293

RESULT 8
S69301
LOCUS S69301
DEFINITION CD34-cell surface antigen {alternatively spliced, clone 2} [mice,
stomal cell line PA-6, mRNA partial, 978 nt].
ACCESSION S69301
VERSION S69301.1
KEYWORDS GI:495715
SOURCE Mus sp. stomal cell line PA-6.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 978)
AUTHORS Suda,J., Sudo,T., Ito,M., Ohno,N., Yamaguchi,Y. and Suda,T.
TITLE Two types of murine CD34 mRNA generated by alternative splicing
JOURNAL Blood 79 (9), 2288-2295 (1992)
MEDLINE 92239883
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 145459] from the original Journal article.
This sequence comes from Fig. 3.
FEATURES
Location/Qualifiers
1..978
/organism="Mus sp."
```

```
gene
1..978
/partial
/gene="CD34"
CDS
1..978
/gene="CD34"
/note="cell surface antigen; This sequence comes from Fig.
4"
/codon_start=1
/product="CD34"
/protein_id="AAB22109.1"
/db_xref="GI:495716"
/translation="MOVHRTDRAGLLLPWRVVALCMLSLHLLNLTSTSTQGIS
PSVTNSESVEENITSSIPGSTSHYLLIQDSSKTPAISETWNTVTSGIPSGSTPH
TFSQPTSPKIGLIPSTSDISITSEMTWKSILPSINVSISPNNSSEFMTSPETPTPT
SSSAPSAIKGKICRGREVLQAGICLSEASCEFEKKEKEDLIIICEKEAE
ADAGASCVSLLAQSEVRPECLMVLNSELPSKLQLMKHKQSLDRKLIGTQSFNQD
IGSHQSYSRKTLIALVTSGVLLAILGTGYFLMNRSSWSPGTGERLELP"
BASE COUNT 261 a 249 c 228 g 240 t
ORIGIN
Alignment Scores:
Pred. No.: 6,76e-51 Length: 978
Score: 887.50 Matches: 192
Percent Similarity: 72.50% Conservative: 40
Best Local Similarity: 60.00% Mismatches: 77
Query Match: 55.78% Indels: 11
DB: 10 Gaps: 4
US-09-836-602-4 (1-316) x S69301 (1-978)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 37 CTGCCATGGCGTGGTAGCTCTGCTGCTGATGAGTCTGCTG-----CAT 81
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 82 CTAATAACTTGCATCTGCTACACGAGAGACTTCTACACAGGAATATCCCATCAGTT 141
Qy 41 SerThrAsnValSerTyrrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 142 CCTACCAATGAGTCTGTTGAGGAAATATACATCTAGCATCCCTGGAAAGTACCAAGCC 201
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrAsnIleThrGluThrThrValLys 80
Db 202 TACTTGATCTATCAGACAGCAGTAGACACACAGCCATCTCAGAGACTATGGTCAAC 261
Qy 81 PheThrSerThrSerValIleThrSerValTyrrGlyAsnThrAsnSerSerValGlnSer 100
Db 262 TTTACAGTTACCTCTGGGATCCCTTCAGGCTCTGGAACTCCACACACTTTTTCACAACCA 321
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 322 CAGACTTCCCCAAGTGGCAGTATGCTGCTACTTCTCAGACAGTATTTCCACTTCAGAGATG 381
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 140
Db 382 ACCTGGAAGTCCAGCCTGCCATCTATAAATGTTCTGATTATTCGCTTAATAATAGCAGC 441
Qy 141 Leu---AlaThrSerProThrLysProTyrr-----ThrSerSerProIleLeu 156
Db 442 TTTGAGATGACATCAACCCACCGCCATATGCTTACACATCATCTCTCTCTCCG----- 495
Qy 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db 496 AGTGCCATTAAAGGAGAAATCAATGCTCTGGAATCCGAGAAGTGAAGTGGCCAGGCT 555
Qy 177 IleCysLeuGlnAlaAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 556 ATCTGCTGCTGAACTAAGTGAAGCATCTAGTTGTGAGGAGTTTAAAGAAAGGAGAGAA 615
Qy 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaGlyAlaGlnVal 216
```

```

Db 616 GATCTAATTCAATACCTGTGTGAAGAGAGAGGCTGAGGCTGATGCTGGTGTAGTGTC 675
Qy 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db 676 TGCCTCCCTGCTTCTAGCCAGCTCTGAGTTAGGCTTCTGAGTCTTGGCTTGGCC 735
Qy 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
Db 736 AATAGCAGACAACCTCCAGCAACTCCAGCTTATGAAAGACCACTATGACTTGAGA 795
Qy 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
Db 796 AAGCTGGGGATCCAAAGCTTCAATAACAAGATATCGGAGCCACCAAGACTATTCCTCGA 855
Qy 277 LysThrLeuLeuAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
Db 856 AAGACTCTTATTGATTTGGTCACTCTGGAGTCTGCTGGCCATCTTGGCCACCACTGGT 915
Qy 297 TyrPheLeuMetAsnArgSerTyrSerProThrGlyGluArgLeuGluLeuPro 316
Db 916 TATTTCCTGATGAACCTCGCAGTTGGAGCCCTACAGGAGAAAGGCTGGAGCTGGAACCT 975

RESULT 9
S69299
LOCUS
DEFINITION
CD34=cell surface antigen (alternatively spliced, clone 6-2-2)
ACCESSION
S69299
VERSION
S69299.1 GI:495713
KEYWORDS
Mus sp. stromal cell line PA-6.
ORGANISM
Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 1149)
Suda,J., Sudo,T., Ito,M., Ohno,N., Yamaguchi,Y. and Suda,T.
Two types of murine CD34 mRNA generated by alternative splicing
Blood 79 (9), 2288-2295 (1992)
92239883
GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 145456] from the original journal article.
This sequence comes from Fig. 3.
Location/Qualifiers
1..1149
/organism="Mus sp."
/db_xref="taxon:10095"
1..1149
/partial
/gene="CD34"
1..1149
/gene="CD34"
/notes="cell surface antigen; This sequence comes from Fig.
3"
/codon_start=1
/product="CD34"
/protein_id="AAB22108.1"
/db_xref="GI:495714"
/translation="MOVHRDTRAGLLPWRVALCMLSLHLNLTSTATTSTOGIS
PSVPTNESVEENITSSIPGSTSHYLIQDSKTPAISTMVNFTVTSIGIPSGCTPH
TSQPOTSPTGILPTTSDSISSEMTWKSLSINVSIDSPNNSSEFMSPPEPYAT
SSAPAIKGLSCSIREVRLAQICLSEASCEEFKKEGDLQILCEKEAE
ADGASVCSLLAQSVREPCLEMLANSTELPSKLQMLKEHQSDRLKLGISFNKD
IGHQSVCSLTIALVTGVLAILGTGTYFLMNRRSWSPGTGERLGEDPYVTENGSGQ
GYSSGPGASPEQGNVHVRGAQENGTGQATSRNHSARQHVVDTEL"

BASE COUNT 307 a 293 c 285 g 264 t
ORIGIN

Alignment Scores:
Pred. No.: 1.13e-49 Length: 1149
Score: 870.50 Matches: 189
Percent Similarity: 71.88% Conservative: 41
Best Local Similarity: 59.06% Mismatches: 79
Query Match: 54.71% Indels: 11

```

```

DB: 10 Gaps: 4
US-09-836-602-4 (1-316) x S69299 (1-1149)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 37 CTGCCATGGCGCTGGTAGCTCTCTGCTGATGAGTCTGCTG-----CAT 81
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 82 CTAATAAATCACTTGACTTCTGCTACCGAGACTTCTACACAAGGAATATCCCCATCAGTT 141
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 142 CCTACCAATGAGTGTGTGAGGAAATATCACATCTAGCATCTCCTGGAAGTACCAGCAC 201
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 202 TACTTGATCTATCAGGACAGCAGTAAGACCAACACCATCTCAGAGACTATGGTCAAC 261
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 262 TTTACAGTTAACTCTGGATCCCTTCAGGCTCTGGAACCTCCACACATTTTCCACACCA 321
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 322 CAGACTTCCCAACTGGCATCTGCTTACTTCTACACAGACTATTTCCTCAGAGATG 381
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 382 ACCTGGAAGTCCAGCTGCCATCTATAAATGTTTCTGATTATTCGCTTAATAATAGCAG 441
Qy 141 Leu---AlaThrSerProThrLysProTyr-----ThrSerSerProIleLeu 156
Db 442 TTTGAGATGACATCACCCACCGAGCCATATGCTTACACATCATCTCTGCTCCG----- 495
Qy 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db 496 AGTGCCATTAAGGAGAGAAATCAAAATGCTCTGGAATCCGAGAAGTGAGTGTGGCCAGGGT 555
Qy 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 556 ATCTGCCCTGGAAGTAAAGTGAAGCATCTAGTTGTGAGAGGTTTAAAGAGGAAAGGGAGAA 615
Qy 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
Db 616 GATCTAATTCAATACCTGTGTGAAGAGAGGCTGAGGCTGATGCTGGTGTAGTGTC 675
Qy 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db 676 TGCCTCCCTGCTTCTAGCCAGCTGCTGAGGTTAGGCCTGAGTGTGCTGTGCTTGGCC 735
Qy 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
Db 736 AATAGCAGACAACCTCCCAAGCTTATGAAAGACCACTATGACTTGTAGG 795
Qy 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
Db 796 AAGCTGGGGATCCAAAGCTTCAATAACAAGATATCGGAGCCACCAAGACTATTCCTCGA 855
Qy 277 LysThrLeuLeuAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
Db 856 AAGACTCTTATTGATTTGGTCACTCTGAGGATCTGCTGGCCATCTTGGCCACCACTGGT 915
Qy 297 TyrPheLeuMetAsnArgSerTyrProThrGlyGluArgLeuGluLeuPro 316
Db 916 TATTTCCTGATGAACCTCGCAGTTGGAGCCCTACAGGAGAAAGGCTGGGTGAAGACCT 975

RESULT 10
S69293
LOCUS
DEFINITION CD34=stem cell antigen [mice, bone marrow, mRNA, 1260 nt].
ACCESSION S69293

```

```

VERSION S69293.1 GI:495708
SOURCE Mus sp. bone marrow.
ORGANISM Mus sp.

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 1260)
Brown, J., Greaves, M.F. and Molgaard, H.V.
The gene encoding the stem cell antigen, CD34, is conserved in
mouse and expressed in haemopoietic progenitor cell lines, brain,
and embryonic fibroblasts
Int. Immunol. 3 (2), 175-184 (1991)
JOURNAL 91223042
MEDLINE GenBank staff at the National Library of Medicine created this
REMARK entry [NCBI gibbsq 145442] from the original journal article.
This sequence comes from Fig. 2.

FEATURES
source Location/Qualifiers
1..1260 /organism="Mus sp."
/db_xref="taxon:10095"
gene 1..1260 /gene="CD34"
CDS 24..1172 /gene="CD34"
/translation="MQVHRTAGLLPWRVWALCLMLLNLNLTATTTSTQGIS
PSVPTNFEVETITSSIPGSTSHYLIYQDSKTPPAISETMNFVTSGIPSGSGTPT
TFSPQSTPGILPTSDISITSEMTWKSILPSINVSDYSPNNSFEMTSPTEPYAVT
SSAPSAIKGICQSGIREVLAOGICLSEASSCEEFKEKEDLIQILCEKEEAE
ADASACVLLLAGVSRPECLMLVLANSTELPSKLOLMKEKHOSDLKIGLQSPNKQD
IGSHQSVSRKTLIALYSGVLLAILGTTGFLMNRRSWSPTEGLGDDPYITNGGGQ
GYSGSGASPTQKQANVTGAENGVTGQATSRNGHSAROHVADTEL"

BASE COUNT 325 a 321 c 320 g 294 t
ORIGIN

Alignment Scores:
Pred. No.: 1.28e-49 Length: 1260
Score: 870.50 Matches: 189
Percent Similarity: 71.98% Conservative: 41
Best local Similarity: 59.06% Mismatches: 79
Query Match: 54.71% Indels: 11
DB: 10 Gaps: 4

US-09-836-602-4 (1-316) x S69293 (1-1260)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 60 CTGCCATGGCGTGGTAGCTCTGCTGCTGAGTCTGCTG-----CAT 104
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 105 CTAAATAACTTCACTCTGCTACCGGAGACTTCTACACAGGAATATCCCATCAGTT 164
Qy 41 SerThrAsnValSerTyGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 165 CTTACCAATAGCTCTGTTCAGGAAATATCACATCATGCTCCCTGGAGTACCGCCAC 224
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 225 TACTTGATCTATCAGACGACGAGTGAAGACCACACCGACCATCTCAGACTATGCTCAAC 284
Qy 81 PheThrSerThrSerValIleThrSerValTyGlyAsnThrAsnSerSerValGlnSer 100
Db 285 TTTACAGTTACCTCTGGGATCCCTCAGGCTCTGGAATCCACACACATTTTTCACACCA 344
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 345 CAGACTTCCCAACTGGGATAGTGGCTACTACTTTCAGACAGATTTTCCACTTCAGAGATG 404

```

```

Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 405 ACCTGGGAAGTCCAGCTGCCATCTATAAATGTTTCTGATTATTCGCCTAAATAAGCAGC 464
Qy 141 Leu---AlaThrSerProThrLysProTyP-----ThrSerSerSerProIleLeu 156
Db 465 TTTGAGATGACATCACCCAGCAGCATATGCTTACACATCATCTCTTCGCTCCG----- 518
Qy 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrClnGly 176
Db 519 AGTGCCATTAAGGGAGAAATCAATGCTCTGGAATCCGAGAGTGAAGTGTGCTCCAGGGT 578
Qy 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 579 ATCTGCTGGAACCTAAGTGAACCATCTAGTTCTGAGGAGTTTAAGAGGAAAGGAGAGAA 638
Qy 197 GlyLeuAlaArgValLeuLysCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
Db 639 GATCTTAATTAATAACTGTTGTGAAAGAGGAGGCTGAGGCTGATGCTGCTAGTCTC 698
Qy 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db 699 TGCTCCCTGCTCTAGCCAGTCTGAGGTAGGCTGAGTGTGCTGCTGCTGCTGCTGCTG 758
Qy 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
Db 759 AATAGCACAGAACTTCCCGACCAAACTCCAGCTTATGAAAGAGCACCACCAATCTGAC 818
Qy 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTySerGln 276
Db 819 AAGCTGGGGATCCAAAGCTTCAATAAACAAGATATCGGAGCCACAGAGCTATTCCCGA 878
Qy 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
Db 879 AAGACTTATTGCTATGCTGCTACCTCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 938
Qy 297 TyrPheLeuMetAsnArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
Db 939 TATTTCTGTGATGAACCGTCCGAGTTGGAGCCCTACAGGAGAAAGCTGGGTGAAGACCT 998

RESULT 11
BC006607 2398 bp mRNA linear ROD 12-JUL-2001
LOCUS Mus musculus, Similar to CD34 antigen, clone MGC:11720
DEFINITION IMAGE:3966337, mRNA, complete cds.
ACCESSION BC006607
VERSION BC006607.1 GI:13879273
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 2398)
Direct Submission
Strausberg, R.
Submitted (27-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalon@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 17 Row: n Column: 8
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES
source

1. .2398

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_MGC="11720 IMAGE:3966337"
/tissue_type="Mammary tumor, Metallothionien-TGF alpha
model, 10 month old virgin mouse, Taken by biopsy."
/clone_lib="NCI_CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
62. 1210
/codon_start=1
/product="similar to CD34 antigen"
/protein_id="AAH06607.1"
/db_xref="GI:13879274"
/translation="MOVHRRTRAGLLLPWRWALCLMSLLHNLTSATETSTQGIS
PSVPTSEVENITSSIPGSHYLIYODSKSTPAISVETWNTVTSIPGSGGTH
TFSPQTSQGTILPTSDSTISSEMTXSLKPSINVDSPNNSSEMTSPTEPYAT
SSAPSA1KBIKCSREVRUAQICIELSPASCEEPKKGEDLIQILCEKBAE
ADAGASVCLLQAQSEVRCAQMLVLANTELPKSLQMKKHQSDLRKLGIOFSNFKQ
IGSHYSRKRTIALVTSGLVLIALLGTTGTFLMNRSSWSPTEGLCEDPVTYENGQ
GYSYSGPASPTQGTNARTRGAEQNTGTGATSRHSHSAROHVVADTEL"

```

CDS

```

62. 1210
/accseq: pcrv_spon10
/codon_start=1
/product="Similar to CD34 antigen"
/protein_id="AAH06607.1"
/db_xref="GI:13879274"
/translation="MOVHRTKAGLLLPWQWVAL
PVSFQSVSEVNTTSIPGSTSHSLFYQDSKPTT
TFSPQSPVPTGILPDTSDTSITSEMTWKSPLP
SSSPASIKLKEIKCSQEVRLVQAIGCLELSEAS
ADAGASVCISLQAGSEVRPECLMLVLANSTELPS
IGSHQSYSRKTLIALVTSQVLLATLGTGTGYFLAN
YSGSPGASSTGCKANVTGKAENGQTQTSRNTS
a 599 c 551 g 625 t

2.91e-49      Length:      2
870.50        Matches:      1
71.06%        Conservative: 4
59.06%        Mismatches: 7
54.71%        Indels:    1
10            Gaps:      4

```

BASE COUNT
ORIGIN

a	599 c	625 t
		551 g
		2398
	Length:	
	Matches:	189
	Conservative:	41
y:	Mismatches:	79
	Indels:	11
	Gaps:	4
		10
	2.91e-49	
	870.50	
	71.68%	
	59.08%	
	54.71%	
	(1-2398)	
	x BC006607	

US-09-836-602-4 (1-316) x BC006607 (1-2398)

Qy	1	MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	98	CTGCCATGGCGCTGGGTAGCTCTGCTGATGATGCTGCTG	142
Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	143	CTAATAACTTGA	202
Qy	41	SerThrAsnValSerTyrrGlnUthrThrThrProSerThrLeuGlySerThrSerLeu	60
Db	203	CCTACCAATCAGTCTGTTGAGGAAATATCACATCTAGCATCCCTGGAGTACCA	262
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys	80
Db	263	TACTTGATCTATCAGGACAGCATAGACACACACGCCATCTCAGAGATATGTGTC	322
Qy	81	PheThrSerThrSerValIleThrSerValTyrrGlyAsnThrAsnSerSerValGlnSer	100
Db	323	TTTACAGTTACCTCTGGGATCCCTTCAGGCTCTGGAACCTCCACACACTTTT	382
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	383	CAGACTTCCCCAATCGGCATCTACTCTACTCTACAGCAGTATTTCCACTTCAGAGATG	442
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	443	ACCTGGAAAGTCCAGCCCGCCATCTAATAATGTTTCTTGATTTTCGGCTAATAATACAGC	502
Qy	141	Leu---AlaThrSerProThrLysProTyr-----ThrSerSerSerProIleLeu	156
Db	503	TTTGAGATGATACATCACCCACCGAGCATATGCTTACACATCATCTTCTGCTCGG	556

Qy	157	SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLysLeuThrGlnGly	176		
Db	557	AGTGCCATTAAAGGAGAAATCAAAATGCTCTGGAAATCGGAGAAGTGAGGTTTGGCCAGGGT	616		
Qy	177	IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu	196		
Db	617	ATCTGCCTGGAACTAAGTGAAGCATCTAGTTGTGAGGAGTTTAAGAAGAAAGGAGAA	676		
Qy	197	GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal	216		
Db	677	GATCTAATTAACAACTACTGTGTAAAGAGGAGGAGCTGAGGCTGATGCTGGTCTAGTGTC	736		
Qy	217	CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla	236		
Db	737	TGCTCCCTGCTTCTAGCCACGCTGAGGTTAGGCTGATGGTCTTGGCTTGGCC	796		
Qy	237	AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys	256		
Db	797	AATAGCAGAGAACTTCCCAAGCAAACTCAGCTTATGAAAGACCACTCTGACTTGAGA	856		
Qy	257	LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln	276		
Db	857	AAGCTGGGGATCCAAAGCTTCAATAAACAGATATCGGAGGCCACAGAGCTATTCCCGA	916		
Qy	277	LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly	296		
Db	917	AAGACTCTTATTTGATTGGTCACCTCTGGAGTTTGTGCTGGCCATCTTTGGGCACACATGGT	976		
Qy	297	TyrPheLeuMetAsnArgSerTyrSerProThrGlyGluArgLeuGluLeuPro	316		
Db	977	TATTTCTGATGAACCTCGCAGTTGGAGCCCTACAGAGAAAGGCTGGGTGAAGACCT	1036		
RESULT 12					
AF202879					
LOCUS	AF202879	581 bp	mRNA linear PRI 02-DEC-1999		
DEFINITION	Homo sapiens hematopoietic progenitor cell antigen CD34 precursor (CD34) mRNA, partial cds.				
ACCESSION	AF202879	GI:6503195			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 581)				
AUTHORS	Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.				
TITLE	Direct isolation and cDNA sequencing of mRNA from human hematopoietic progenitor cell antigen CD34				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 581)				
AUTHORS	Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-NOV-1999), Medicine, Hematology and Oncology, Sickie Cell Center, Medical College of Georgia, 15th Street, AC-1000, Augusta, GA 30912, USA				
FEATURES	Location/Qualifiers				
source	1..581				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="1"				
	/map="1q32"				
gene	<1..>581				
	/gene="CD34"				
CDS	<1..>581				
	/gene="CD34"				
	/codon_start=2				
	/product="hematopoietic progenitor cell antigen CD34 precursor"				
	/protein_id="AAFI4634.1"				
	/db_xref="GI:6503196"				
ARVLCPEQADADAGACVCSLLAOSREVPQCLTKTGTCLEQNKTSSTCAFFKKDKRGGL /translation="ILSDIKAEIKCSGIREVKLTQGTCLQNKTSSTCAFFKKDKRGGL ARVLCPEQADADAGACVCSLLAOSREVPQCLTKTGTCLEQNKTSSTCAFFKKDKRGGL					

```

/translation="ILSDIKAEIKCSGIREVKLTQICLEQNKTSSCAEFKKDRGEG
ARVLGEECADADAGAOVCSLLAOSEVRPOCLLVLANRTEISSKLOLMKKHOSDLK

```


Db	605	GTGAAATGGCCCAACAGCATCTGCCTGAAGCTAATAATGAGACTTCAGCTGTGAGGAGTTT							
Qy	191	LysLysAspArggLyGluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAla							
Db	665	AAGAAGAACAAATGAAGAGAAACTGACCGGATTCTGTGTCAGAGAAACAG-----							
Qy	211	AspAlaGlyAlaGlnValCysSerLeuLeuAlaGlnSerGluValAlaArgProGlnCys							
Db	716	-----GCCGAGGTGTCTCTGCTGCTGCCCACTGAGTGAGGTCAGGCTCAGTGC							
Qy	231	LeuLeuLeuValLeuAlaAsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLys							
Db	767	CTGCTGCTGGTCTGACCACACAGACAGAAAGTCCACACAGATCAAACCTCTGGAGAG							
Qy	251	HisGlnSerAspLeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValalaSer							
Db	827	CAC TG GGT GTG ACCT CG AG A GA T GG C AT C C G G C C A C T C T C T G A A G A A G C T C T C A G C A G C							
Qy	271	HisGlnSerTy rSerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuAla							
Db	887	CAC C A G A C T A T T C C C G G A A G A C C T T G A T T C A C T G T C A C C T C G G G A T C C T G C T G C C							
Qy	291	ValLeuGlyIleThrGlyTy rPheLeuMetAsnArgSerTrpSerProThrGlyGlu							
Db	947	A T C T T G A T C A C C A C G T G C T A T T T C C T G A T G A A C C C G C C A C T G G A G C C C A A C G G A G A A							
Qy	311	ArgLeuGluLeuGluPro	316						
Db	1007	AGGCTGGCGGAAGACCCCT	1024						
RESULT	14								
LOCUS	AF259378	Capra hircus CD34 mRNA, partial cds.	2420 bp	mRNA	linear	MAM	14-JUN-2000		
DEFINITION	AF259378								
ACCESSION	AF259378								
VERSION	AF259378.1	GI:8515725							
KEYWORDS									
SOURCE	goat.								
ORGANISM	Capra hircus								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;								
TITLE	Bovidae; Caprinae; Capra.								
JOURNAL	1 (bases 1 to 2420)								
REFERENCE	Leipprandt,J.R., Anson,D.S. and Jones,M.Z.								
AUTHORS	Cloning and sequence analysis of caprine CD34 cDNA								
TITLE	Unpublished								
JOURNAL	2 (bases 1 to 2420)								
REFERENCE	Leipprandt,J.R., Anson,D.S. and Jones,M.Z.								
AUTHORS	Direct Submission								
TITLE	Submitted (20-APR-2000) Biochemistry, Michigan State University,								
JOURNAL	Room 519 Biochemistry, East Lansing, MI 48824, USA								
FEATURES	Location/Qualifiers								
source	1..2420								
	/organism="Capra hircus"								
	/db_xref="taxon:9925"								
CDs	<1..1069								
	/note="cell-surface antigen"								
	/codon_start=2								
	/product="CD34"								
	/protein_id="AAF76145.1"								
	/db_xref="GI:8515726"								
	/translation="PSMEFTNTVATNGATVSSNPGISSPVPTTTSYWKSMNSSTFRST								
	SLYNVDSDNGTAVTSVPASTVNFTSASETLPASGVANSVSQPOTSLATASSTLIS								
	FATSEALTOPSTPGNISLYNSTSPVTSSINTSPSFPTQNILKEIKCSRVKVK								
	LAHSICLKLNETSCEEFKNNEKLTVLCQEQAAYCVCSKLLAQAEVKPQCILLVL								
	NRTESSKRLKARYQDLREMGIDISEVSSHOSYSRKLTIALVTSGILAILALT								
	TCYFLMNRHSSPAERLGEDPTFYVNGGQQYNSGSEAPGAQGKAIVNRGAQENG								
	GQATSRHGSARQRVADTEL"								
BASE COUNT	565 a 700 c 590 g 565 t								
ORIGIN									

Pred. No. :	1.21e-40	Length:	2420
Score:	741.50	Matches:	163
Percent Similarity:	68.37%	Conservative:	38
Best Local Similarity:	55.44%	Mismatches:	84
Query Match:	46.61%	Indels:	9
DB:	4	Gaps:	3
US-09-836-602-4 (1-316) x AF259378 (1-2420)			
Qy	25	GlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnVal	44
Db	35	GGTAATGCCACAGTGTCA---TCCAACCCAGGATATCTTCACCTGTCCCTACAACTACG	91
Qy	45	SerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSer	64
Db	92	TCCCTACTGGAAGAAGCATCAACTCAAGTACTTTTGAAGACACCAGCCTCTCAATGTCTCT	151
Qy	65	Gln-----HisGlyAsnGluAlaThrThraAsnIleThrGluThrThrValLysPheThr	82
Db	152	CAGGACAGCAATGGGACACAGTAGTCACCTCAGCTCCAGCGTCTACAGTCAATTTTCACG	211
Qy	83	SerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGlnThr	102
Db	212	TCTGCCTCTGAGACCCCTCCAGCCCTCTGGAGCCGTGAACCTCTCTGTCCAACACAGACC	271
Qy	103	SerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrThrLeu	122
Db	272	TCTTTAGCCACACGGCATCTTCTACCCCTCATCAGCTTTGCGACTTCAGAGCGGACTCTG	331
Qy	123	LysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSerLeuAla	142
Db	332	CAGCCACAGCAGTTCTCTGGAAATATTTCGGATTCCCTATACAAATAGTACCAAGCCCTGTG	391
Qy	143	ThrSerProThrLysProTyrThrSerSerSerProfileLeuSerAspIleLysAlaGlu	162
Db	392	ACTTCTTCCATTAAATACCTCTCCATCATTTTCTCCTACCCCAAAATATCCTCAAGACTGAA	451
Qy	163	IleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGluGlnAsn	182
Db	452	ATAAAATGTTCCAGAGTCAAGAAAGTGAATTTGCCCCACAGCATCTCGCTGAAGCTAAAT	511
Qy	193	LysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArgValLeu	202
Db	512	GAGACCTCCAACTGTGAGGAGTTTAAAGAAGCAATGAAGAGAGCTGACCCGAGTCCCTG	571
Qy	203	CysGlyGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeuLeuAla	222
Db	572	TGTCAGGAGGAACAG-----GCCAGGTGTCTCTCTGCTCTCTGCTCTGCTCTGCTCTG	613
Qy	223	GlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGluIleSer	242
Db	614	CAGTCTGAGGTGAAGCCTCAGTGCCTGCTGTGCTCTGACCAACAGGACAGAAAGTTCC	673
Qy	243	SerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIleLeuAsp	262
Db	674	AGCAAGATCAGACTTCTGAAAGGTACACGTCTGACCTGAGAGAGATGGCCTCCAGGAC	733
Qy	263	PheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIleAlaLeu	282
Db	734	ATCTCTGAAGAAGATGTACAGCGCCACAGAGCTATTTCGCGAAGACCTTGTATTGCACTG	793
Qy	283	ValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMetAsnArg	302
Db	794	GTCACCTCGGGATCTCTGCTGGCCATCTCATCACCACGTGCTATTTCCTTGATGAACCGC	853
Qy	303	ArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro	316
Db	854	CGCAGCTGGAGCCCCAGCAGGAGAAAGGCTGGGCGAAGACCCCT	895
RESULT 15			
HS88L2/c			
LOCUS			
DEFINITION Human DNA sequence from clone 812 on chromosome 1q32.2-q32.3			
160771 bp DNA linear PRI 22-NOV			

Alignment Scores:

Contains CD34 (hematopoietic stem cell surface antigen), CA repeat, EST, GSS, complete sequence.

ACCESSION
AL035091
VERSION
KEYWORDS
SOURCE

2 GI:4581417
HTG; CD34; repeat polymorphism.
human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160771)
Graham.D.

REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Submitted (22-FEB-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

On Apr 12, 1999 this sequence version replaced gi:4140364.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 88L2.
The true right end of clone 7H11 is at 71906 in this sequence. 8L2 is from the library RCI6 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR.pPAC4>.

FEATURES

Source

1. .160771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q32.2-q32.3"
/clone="XX-D88L2"
/clone.lib="RPC16"

repeat_region

/note="10 copies 4 mer aaga 83% conserved"

repeat_region

42. .75
/note="17 copies 2 mer aa 88% conserved"

repeat_region

89. .412
/note="LIMC5 repeat: matches 7609. .7913 of consensus"

repeat_region

1027. .1154

/note="L2 repeat: matches 2586. .2744 of consensus"

repeat_region

1189. .1277

/note="L2 repeat: matches 2616. .2704 of consensus"

repeat_region

2426. .2643

/note="MER58A repeat: matches 1. .224 of consensus"

misc_feature

complement(3075..3399)

repeat_region

/note="match: 252805 STR containing (CA) repeat"

repeat_region

3121. .3176

/note="28 copies 2 mer gt 100% conserved; differs from 252805"

repeat_region

3806. .3905

/note="L2 repeat: matches 2600. .2750 of consensus"

repeat_region

8052. .8339

/note="L2 repeat: matches 1823. .2111 of consensus"

repeat_region

8406. .8600

/note="MLT1A1 repeat: matches 162. .365 of consensus"

repeat_region

8631. .8815

/note="L2 repeat: matches 2158. .2346 of consensus"

repeat_region

8843. .9132

/note="AluJb repeat: matches 1. .292 of consensus"

repeat_region

9295. .9363

/note="L2 repeat: matches 2680. .2749 of consensus"

9372. .9672

/note="match: GSS AQ075766 clone 2359015"

9493. .9594

/note="MIR repeat: matches 96. .205 of consensus"

9619. .9640

/note="11 copies 2 mer tc 100% conserved"

9671. .9835

/note="MIR repeat: matches 46. .200 of consensus"

10253. .10285

/note="L2 repeat: matches 2708. .2740 of consensus"

10284. .10406

/note="MIR repeat: matches 64. .196 of consensus"

<11347..11584

/note="match: multiple ESTs; match: AA884400 AA971811"

11846. .11878

/note="MIR repeat: matches 1. .33 of consensus"

11879. .12182

/note="AluJ repeat: matches 1. .303 of consensus"

12183. .12411

/note="MIR repeat: matches 33. .258 of consensus"

12443. .12491

/note="L2 repeat: matches 2662. .2710 of consensus"

12574. .12661

/note="MIR repeat: matches 178. .262 of consensus"

12756. .13049

/note="AluJb repeat: matches 1. .300 of consensus"

13231. .13541

/note="match: STS G28100, G24984"

13231. .151597

/note="match: multiple ESTs; match: 3' EST T89361 clone 110082; Paired with EST T84126 matching this clone; match: 5' EST T84126 clone 110082; Paired with EST T89361 matching this clone"

13672. .13875

/note="MIR repeat: matches 29. .231 of consensus"

14598. .14907

/note="AluJ repeat: matches 1. .310 of consensus"

14913. .15214

/note="AluJb repeat: matches 7. .294 of consensus"

15376. .15596

/note="L2 repeat: matches 1263. .1527 of consensus"

15606. .15666

/note="MLT1F repeat: matches 478. .541 of consensus"

15667. .16045

/note="L1P43 repeat: matches 5764. .6146 of consensus"

16046. .16084

/note="MLT1F repeat: matches 438. .478 of consensus"

16085. .16379

/note="AluJb repeat: matches 1. .296 of consensus"

16380. .16850

/note="MLT1F repeat: matches 1. .438 of consensus"

16854. .17208

/note="L2 repeat: matches 1533. .1930 of consensus"

17268. .17395

/note="L1M4 repeat: matches 7854. .7977 of consensus"

17422. .17485

/note="Charliel repeat: matches 79. .143 of consensus"

17509. .17625

/note="L1M4 repeat: matches 7706. .7822 of consensus"

17627. .17677

/note="L2 repeat: matches 1949. .1995 of consensus"

17678. .17988

/note="AluJ repeat: matches 1. .311 of consensus"

17989. .18227

/note="L2 repeat: matches 1995. .2257 of consensus"

18250. .18508

/note="Charliel5 repeat: matches 2315. .2577 of consensus"

18510. .18698

/note="Charliel5 repeat: matches 16. .196 of consensus"

18721. .18935

/note="Charlielb repeat: matches 3. .219 of consensus"

18936. .19243

```
/note="AluY repeat: matches 1. .307 of consensus"
19244. .19531
/note="Charlieb repeat: matches 219. .501 of consensus"
19610. .19779
/note="L1MD repeat: matches 1. .169 of consensus"
19780. .20228
/note="L1MD repeat: matches 787. .1293 of consensus"
20299. .20606
/note="L2 repeat: matches 2382. .2710 of consensus"
20625. .20753
/note="L2 repeat: matches 2179. .2307 of consensus"
20813. .20960
/note="L2 repeat: matches 2374. .2523 of consensus"
21149. .21292
/note="MIR repeat: matches 110. .261 of consensus"
21334. .22002
/note="L2 repeat: matches 1947. .2724 of consensus"
<21938. .22360
/note="match: GSS AQ193501 clone 2383C13"
<22003. .>22108
/note="match: multiple ESTs; match: AA909597 AA906144
AA907722"
22573. .22657
/note="MIR repeat: matches 81. .160 of consensus"
complement(24084. .25666)
/note="match: STS G06650"
27028. .27184.35057. .35137.36529. .36782.37377. .37559,
48563. .48899))
/gene="CD34"
/note="match: M81104; x60172; match: multiple ESTs; match:
T80274 C18678 AA553071 AA311526 AA311473 AA907722 AA906144
AA064307 A1173145 AA369014 R39493 AA148360 T81557 R72478
R69703 R69740 AA039660 RA639555 R73090 R73955 W92532
R66845 AA054965 AA434483 R69214 R82715 AA188435 AA524276
W74409 H01096 T12007 AA216274 AA022916 AA528783 AA483672
AA640408 W58493 T28504 A1128488 A1149563 A1144193 A1252665
A1017002 A1160509 A1017793 AA906021 A1128776 A1150834
AA434387 AA664247 AA039661 W72884 A1262206"
/evidence="not_experimental"
complement(24102. .48899)
/gene="CD34"
CDS
complement(join(25301. .25486.26245. .26409.26724. .26776,
27028. .27184.35057. .35137.36529. .36782.37377. .37559,
48563. .48605))
/gene="CD34"
/note="match: SWISS-PROT:P28906"
/codon_start=1
/evidence="not_experimental"
/product="CD34 (hematopoietic stem cell surface antigen)"
/protein_id="CAB36556.1"
/db_xref="GI:4455631"
```

Alignment Scores:

```
Pred. No.: 5,35e-30 Length: 160771
Score: 617.00 Matches: 145
Percent Similarity: 42.32% Conservativity: 1
Best Local Similarity: 42.03% Mismatches: 1
Query Match: 38.78% Indels: 198
DB: 9 Gaps: 1

US-09-836-602-4 (1-316) x HS88L2 (1-160771)

QY 14 LeuProSerGlyPheMetSerLeuAsnAsnGlyThrAlaThrProGluLeuProThr 33
Db 37563 ATAGCTTCGGGTTTCATGAGTCTTGACAAACACGGTACTGCTACCCCGAGATTACTACC 37504

QY 34 GlnGlyThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrProSer 53
Db 37503 CAGGGAACATTTTCAATGTTTCTACAAATGATCTACCAAGAAACTACACACCTAGT 37444

QY 54 ThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsn 73
|||||
```

Search completed: October 30, 2002, 09:52:02
Job time : 1666.9 secs

```
Db 37443 ACCCTTGGGAAGTACCAGCCTGCACCCCTGTGTCTCAACATGGCAATGAGCCACACAAC 37384
QY 74 ILeThr----- 75
|||||
Db 37383 ATCACAGGTAAACACAGCATTTTGTGCAGATCCCGGAGAGATGCTGGTGATGCTTGGGTA 37324
QY 75 ----- 75
Db 37323 AAGCATTTAGGATGTTTTCAGACCGCTCCCTCTCCACAGAGAAATTTATACAAGTCCCT 37264
QY 75 ----- 75
Db 37263 AGTATTAAATGACTTGGAGTATCATGTTAGGGTGCCCTGAAGTAAGTCTTAGATAATCTTT 37204
QY 75 ----- 75
Db 37203 CCTTCAGTCACAGTTTCTTCAGGCCCATGTCTTGGGAGCTGATCCTCGATCAGTAATGCCT 37144
QY 75 ----- 75
Db 37143 CCATACCCCTTCTCTCATGTGAGTCTAGTCTAGTATAAAATAAGGATCAGAAGACTCA 37084
QY 75 ----- 75
Db 37083 ATAGAGATCTCCGATCTCTTCTTAAAGAAAGAGGTGGGTACAGGTCCAAACTGGGGTC 37024
QY 75 ----- 75
Db 37023 TTTGGCTTCTCTAAGTAGACACCATCTATTTCAGTTTCAGGACACCTCTTAATACATTC 36964
QY 75 ----- 75
Db 36963 TAGATTCTAGCTCTGTGTCCACAGCTTGAAATGAGTTTGGTCAGGATGGGACACGAAG 36904
QY 75 ----- 75
Db 36903 TAACTGTTAACTCCCAAACTCCCTTTCTGTGTGTTGAAGTGCATGCCATATCATGACCT 36844
QY 75 ----- 75
Db 36843 GAGATTTTGTGTATCTATGACGACATGCACACATGCACACCATGTTTGGTCTCTTCCA 36784
QY 76 GluThrThrValLysPheThrSerThrSerValIleThrSerValIleThrSerValIleThrAsn 95
|||||
Db 36783 GAAACGACAGTCAAATTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACACAAC 36724
QY 96 SerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnVal 115
|||||
Db 36723 TCTTCTGTCCAGTCACAGACCTCTGTAATFCAGCAGAGTTTCACCCACCCACCCACGTT 36664
QY 116 SerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSer 135
|||||
Db 36663 TCACTCCAGACACAACCTTGAGGCTAGCCCTGTACCTGGGAATGTTTTCACACCTTCA 36604
QY 136 ThrThrSerThrSerLeuAlaThrSerProThrLysProTyrThrSerSerSerProIle 155
|||||
Db 36603 ACCACTAGCAGTACCTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTCTATC 36544
QY 156 LeuSerAspIleLys 160
|||||
Db 36543 CTAAGTGACATCAAG 36529
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:10:53 ; Search time 1294.57 seconds

(without alignments)
3888.820 Million cell updates/sec

Title: US-09-836-602-2

Perfect score: 1895

Sequence: 1 MPRGWTALCLLLPSGFMS.....QATSRNGHSARQHVADTEL 373

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlip
-Q=/cgn2_1/USPto_Spool/US09836602/runat_29102002_101144_3812/app_query.fasta_1.1429
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09836602_SGNN_1_1_2453_@runat_29102002_101144_3812 -NCPU=6 -ICPU=3
-NO_XLPIP -NO_MMIP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1548	81.7	977	9	AL548050
2	1409	74.4	919	9	AL513812

RESULT 1

AL548050

LOCUS

DEFINITION

AL548050

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

ALIGNMENTS

AL548050 977 bp mRNA linear EST 16-FEB-2001
AL548050 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1034YP22 5
prime, mRNA sequence.
AL548050.1 GI:12882692
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 977)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1034YP22"
/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 259 a 272 c 231 g 214 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 5.18e-123 Length: 977
Score: 1548.00 Matches: 312
Percent Similarity: 99.68% Conservative: 0
Best Local Similarity: 99.68% Mismatches: 1
Query Match: 81.69% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AL548050 (1-977)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 40 ATGCCGGGGCTGGAGCGCGCTTGTGCTGAGTTGCTGCTTCTGGGTTTCATGAGT 99
QY 21 LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 100 CTTGACAAACAGGTACTGTACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 159
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 160 TCTACAATGTATCTACCAAGAAACTACACACCTAGTACCTTGGAGTACAGCGCTG 219
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 220 CACCCTGTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAAGACAGTCAAA 279
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 280 TTCACATCTACCTCTGATGAATCACTAGTTATGGAAACAAACATCTTCTGTCAGTCA 339
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 340 CACACCTCTGTATACACACAGTGTTCACCCACCCAGCCCAACGTTTCACTCCAGAGCA 399
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 400 ACCTTGAAGCCTAGCCCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCACTAGC 459
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
Db 460 CTTGCAACATCTCCCACTAAACCTATACATCATCTCTCTATCTCAAGTGACATCAAG 519
QY 161 AlaGluIleLysCysSerGlyIleAtrGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 520 GCAGAAATCAAAATGTTTCAGGCATCAGAGAACTGAATGTACTCAGGGCATCTCCCTGGAG 579
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 580 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGGGGAGGCGCTGGCCCGA 639
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 640 GTCTGTGTGGGAGAGCAGGCTGATGCTGATGCTGGGGCCAGGATGCTCCCTGCTC 699
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 700 CTTGCCCCAGTCTCAGGTGAGGCCCTCAGTGTCTACTGCTGTGCTTGGCCAAACAGAGAA 759
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260

Db 760 ATTTCCAGCAATCCCAACTTATGAAAAGCACCACAACTCTGACCTGAAAAGAGCTGGGATC 819
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 820 CTGATTTTCACTAGCAGAGATGTTGCAAGCCACAGAGCTATTCCCAAGAGACCTGATT 879
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 880 GCATGCTACCTCGGAGGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCCTCATG 939
QY 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGly 313
Db 940 AATCGCGCAGCTGGAGCCCMCA-GGAGAAAGGCTGGGG 977

RESULT 2

AL513812 919 bp mRNA linear EST 13-FEB-2001
LOCUS AL513812 LTI_NFL006_PL2 Homo sapiens cDNA clone CL0BA0092F10 5
DEFINITION prime, mRNA sequence.

ACCESSION AL513812
VERSION AL513812.1 GI:12777306
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 919)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqrefegenoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

Location/Qualifiers

1..919

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CL0BA0092F10"

/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"

/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 246 a 253 c 213 g 198 t 9 others

ORIGIN

Alignment Scores:

Pred. No.: 4.22e-111 Length: 919

Score: 1409.00 Matches: 286

Percent Similarity: 97.61% Conservative: 0

Best Local Similarity: 97.61% Mismatches: 7

Query Match: 74.35% Indels: 1

DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AL513812 (1-919)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20

Db 40 ATGCCGGGGCTGGAGCGCGCTTGTGCTGAGTTGCTGCTTCTGGGTTTCATGAGT 99

QY 21 LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

Db 100 CTTGACAAACAGGTACTGCKACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 159

```
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 160 TCTACAAATGATCCYACCAAGAACTACACACCTAGTACCTTGGAGTACCAGCTG 219
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVallys 80
Db 220 CACCTGTGTCTCAACATGCAATGAGCCACACAAACATCACAGAAACGACAGTAA 279
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 280 TTCACATCTACCTCTGTGTAACCTCAGTTATGGAACACAACTCTCTGTCCAGTCA 339
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 340 CAGACCTCTGTATACACACAGTGTTCACACCCGCCACCGTTCACCTCCAGACACA 399
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 400 ACCCTGAAGCCTAGCCTGTACCTGGAAATGTTTCAGACCTTTTCAACCACTAGC 459
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 460 CTTCACATCTCCCTCACTAACCTATACATCATCTCTCTATCCYAAGTGACATCAAG 519
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 520 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATGACTCAGGGCATCTGCCTGGAG 579
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 580 CAAATAAGACCTCCACCTGTGCGGAGTTTAAAGAGACAGGGGAGAGGCGCTGGCCGA 639
Qy 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 640 GTGCTGTGTGGGAGAGAGAGGCTGATGCTGCTGGGCCAGGATGCTCCCTGCTC 699
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
Db 700 CTTGCCAGTCTGAGGTGAGGCTCAGTGTCTACTGTGCTGTTGGCCACAGAACAGAA 759
Qy 241 IleSerSerLysLeuGlnLeuMetLysHisGlnSerAspLeuLysLysLeuGlyLe 260
Db 760 ATTTCCAGCAAACTCCAACTTATGAAAGACCACTGACCTGAAAGAGCTGGGATC 819
Qy 261 LeuAspPheThrGlnGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 820 YTAGATTTCAGTGAGCAGAGAGTGTGCAAGCCACAGAGCTATTTCCCAAGACCCCTGATT 879
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGly 293
Db 880 GCACCTGGTCACTCGGAGC-CTGCTGGTGTCTTTGGGC 917

RESULT 3
ALS98821 735 bp mRNA linear EST 14-AUG-2001
LOCUS DKFZP313K2322_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone
DEFINITION DKFZP313K2322 5', mRNA sequence.
ACCESSION ALS98821
VERSION ALS98821.1 GI:15161512
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 735)
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
,S.
TITLE EST (Duesterhoeft, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Duesterhoeft A
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
```

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp313K2322) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source
1. .735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp313K2322"
/clone_lib="313 (synonym: hlcc2)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"
BASE COUNT 191 a 205 c 173 g 164 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 3.1e-89 Length: 735
Score: 1153.00 Matches: 233
Percent Similarity: 98.31% Conservative: 0
Best Local Similarity: 98.31% Mismatches: 2
Query Match: 60.84% Indels: 2
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x ALS98821 (1-735)

Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 28 ATGCCGGGGGCTGGACCGCCTTGTGCTGTGAGTTGCTGCCTTCGGGTTCTAGAGT 87
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProGlnGlnGlyThrPheSerAsnVal 40
Db 88 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAATGTT 147
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 148 TCTCAAAATGATCTCTACCAAGAACTACAACACTAGTACCTTGGAGTACCAGCTG 207
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 208 CACCTGTGTCTCAACATGGCAATGAGCCACACAAACATCACAGAAACGACAGTCAA 267
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 268 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTCTGTCCAGTCA 327
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 328 CAGACCTCTGTAAATCAGACACAGTGTTCACACCCCGCAGCAACGTTTCAACTCCAGACACA 387
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 388 ACCCTGAAGCCTAGCCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCAGTAC 447
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 448 CTTGCAACATCTCCCACTAAACCTATACATCATCTCTCTCTATCTTAAAGTGACATCAAG 507
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 508 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATGACTCAGGGCATCTGCCTGGAG 567
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLys-LysAspArg-GlyGluGlyLeuAlaA 200
Db 568 CAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAAAGACAGGGGAGAGGCGCTGGCCC 627
Qy 200 rgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuL 220
Db 220

Db	628	GAGTGTGTGTGGGGAGGAGCAGCGCTGATGCTGATGCTGTGGGGCCCGCAGGTATGCTNCCTGC	687
Qy	220	euLeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeu	235
Db	688	TCCTTGCCAGTCTGAGGNGAGGCCTCAGTGTCTACTGCTGTGCTTG	734
RESULT 4			
AU141253			
LOCUS	AU141253	THYR01 Homo sapiens cDNA clone	761 bp mRNA linear EST 25-OCT-2000
DEFINITION	AU141253	sequence.	
ACCESSION	AU141253		
VERSION	AU141253.1	GI:11002774	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 761)		
	Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,		
	Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and		
	Isogai,T.		
TITLE	HRI human cDNA project		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Takao Isogai		
	Genomics Laboratory		
	Helix Research Institute		
	1532-3 Yana, Kisarazu, Chiba 292-0812, Japan		
	Tel: 81-438-52-3951		
	Fax: 81-438-52-3952		
	Email: genomics@hri.co.jp		
	HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix		
	Research Institute; cDNA library construction: Department of		
	Virology, Institute of Medical Science, University of Tokyo, and		
	Helix Research Institute.		
FEATURES	Location/Qualifiers		
source	1..761		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="HYR01000286"		
	/clone_lib="THYR01"		
	/tissue_type="thyroid gland"		
	/note="Vector: pME18SFL3"		
BASE COUNT	197 a 214 c 180 g 167 t		3 others
ORIGIN			
Alignment Scores:			
Pred. No.:	1.51e-86	Length:	761
Score:	1122.00	Matches:	222
Percent Similarity:	98.68%	Conservative:	2
Best Local Similarity:	97.80%	Mismatches:	3
Query Match:	59.21%	Indels:	0
DB:	9	Gaps:	0
US-09-836-602-2 (1-373) x AU141253 (1-761)			
Qy	1	MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	73	ATGCCGGGGGCTGGACCGCGCTTTGCTGTGAGTTTGTGCGCTTCCTGGGTTTCATGAGT	132
Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	133	CTTGACAACAACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAATGTT	192
Qy	41	SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu	60
Db	193	TCTACAATGTATCTCTACCAAGAACTACAAACACCTAGTACCCCTTGGAAAGTACCAGCTG	252
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValIys	80
Db	253	CACCCTGTGCTCAACATGGCAATGAGGCCACAACAAGATCACAGAACACAGTCAAA	312
Qy	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100

Score: 1096.00 Matches: 218
Percent Similarity: 98.20% Conservative: 0
Best Local Similarity: 98.20% Mismatches: 4
Query Match: 57.84% Indels: 0
Db: 9 Gaps: 0

US-09-836-602-2 (1-373) x AU141111 (1-752)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 65 ATGCGCGGGCTGACCGCGCTTGTGCTGAGTTGCTGCCCTTCGGGTTTCATGAGT 124

Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 125 CTTGACAAACAGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTCAATGTT 184

Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 185 TCTACAAATGATCTCTACCAAGAACTACACACCTAGTACCTTTGGAGTACCAAGCCTG 244

Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 245 CACCCTGTGCTCAACATGGCAATGAGGCCCAACAACATCACAGAACGACGTCAA 304

Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 305 TTCACATCTACCTCTGTGATACCTCAGTTTATGAAACACAACTCTCTGCCAGTCA 364

Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 365 CAGACCTCTGTAATCAGCACAGTGTACACCCAGCCCAACGTTTCAACTCCAGAGACA 424

Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 425 ACCTTGAGCCCTAGCCTGTACCTCGAAATGTTTCAGACCTTTCACCACTAGCAGTAC 484

Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 485 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTCTATCTTAAGTGACATCA 544

Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyLysCysLeuGlu 180
Db 545 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCCG 604

Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysAspArgGlyClyGluLysLeuAlaArg 200
Db 605 CAAATAAGACCTTCAGCTGCGGAGTTTAAGAAGCAGGGGAGAGGGCCTGGCCCCGA 664

Qy 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 665 GTGCTGTGGGAGGAGCAGCTGATGCTGATGCTGGGGCCCAAGTATGCTCTGCTC 724

Qy 221 LeuAla 222
Db 725 CTTGCC 730

RESULT 6
LOCUS AU121093
DEFINITION AU121093 HEMBB1 Homo sapiens cDNA clone HEMBB1002065 5', mRNA
sequence.
ACCESSION AU121093
VERSION AU121093.1 GI:10936328
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 749)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project

JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES
source
1..749
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBB1002065"
/tissue_type="whole embryo, mainly body"
/dev_stage="embryo, 10 weeks"
/note="Vector: pME18SFL3"

BASE COUNT 195 a 211 c 173 g 166 t
ORIGIN

Alignment Scores:
Pred. No.: 8.93e-83 Length: 749
Score: 1078.00 Matches: 214
Percent Similarity: 96.83% Conservative: 0
Best Local Similarity: 96.83% Mismatches: 7
Query Match: 56.89% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AU121093 (1-749)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 73 ATGCGCGGGCTGACCGCGCTTGTGCTGAGTTGCTGCCCTTCGGTTCATGAGT 132

Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 133 CTTGACAAACAGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTCAAAATGTT 192

Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 193 TCTACAAATGATCTCTACCAAGAACTACACACCTAGTACCTTGGAGTAGTACCACCTG 252

Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 253 CACCCTGTGCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAA 312

Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 313 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCAGTCA 372

Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 373 CAGACCTCTGTAATCAGCACAGTGTCCACCCAGCCAGCAACGTTTCACTCCAGAGACA 432

Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 433 ACCTTTGAAGCCTTAGCCGCTCACCTGGAAATGTTTTCAGACCTTTCAACCACTAGCAGTAC 492

Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 493 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTATCTTAAGTACATCAAG 552

Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyLysCysLeuGlu 180
Db 553 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGCGCTGCCCTGGAG 612

Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 613 CAAATAAGACCTTCAGCTGTGCGGAGTTTAAGAAGCAGGGGAGGANGSCCTGGCCCCA 672

```

QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 673 GTGCTGTGGGGAGGAGACANGCTGATGCTGATGCTGGGGCCCAAGTATGCTCTGCTCC 732
QY 221 Leu 221
|||
Db 733 TTG 735

RESULT 7
AUI41883
LOCUS AUI41883 THYR01 Homo sapiens cDNA clone THYR01001399 5', mRNA
DEFINITION sequence.
ACCESSION AUI41883
VERSION AUI41883.1 GI:11003404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS Oca,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamanoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01001399"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/notes="Vector: pME18SFL3"
BASE COUNT 198 a 215 c 181 g 171 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 4.98e-82 Length: 768
Score: 1069.50 Matches: 217
Percent Similarity: 95.61% Conservative: 1
Best Local Similarity: 95.18% Mismatches: 10
Query Match: 56.44% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AUI41883 (1-768)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 72 ATGCCGGGGGCTGGACCGGCTTTGCTGCTGAGTTTGCTGCTTGGGTTTCATGAGT 131
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 132 CTGTGACAAACACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTCAATGTT 191
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 192 TCTACAAATCATCTACCAAGAACTACAACTAGTACCTTGGAGTACGAGCTG 251
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrThrAsnIleThrGluThrValLys 80
|||||

```

```

Db 252 CACCTGTGCTCAACATGGCAATGAGCCACACAACAACACACAGAAACACAGTCAAA 311
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 312 TTCATCTCTACCTCTGTGATAACCTCAGTTATGGAACACAAAACCTCTCTGTCAGTCA 371
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 372 CAGACCTCTGTAATCAGCACAGTGTTCACCACGCCAGCCAGCGTTTCAACTCCAGAGACA 431
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 432 ACCTTGAAGCTAGCTACCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCACTAGC 491
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 492 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCCTAAGTGACATCAAG 551
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 552 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAAATGACTCAGGGCATCTGCCTGGAG 611
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 612 CAAAATAAGACCTTCAGCTGTGGGAGTTTAAAGAAGACAGAGGGAGGCGCTGGCCCGA 671
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 672 GTGCTGTGGGGAGGAACANGCTGATGCTGATGCTGGGC--CCAGTATGCTTCTCTCTC 729
QY 221 LeuAlaGlnSerGluValArgPro 228
|||||
Db 730 CTTGCCAATCTGANGTGAAGGCCT 753

RESULT 8
BI769224
LOCUS BI769224
DEFINITION mRNA sequence.
ACCESSION BI769224
VERSION BI769224.1 GI:15760802
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1129)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11525 row: e column: 14
High quality sequence stop: 765.
FEATURES
Location/Qualifiers
1..1129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5209333"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/notes="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site:1: NotI; Site:2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed

```

Qy 1 MetProArgGlyTrrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 46 ATGCCCGGGGCTGGACCGGCTTTGGTGTGTAGTTTGTGCTTCTGGGTTTCATGAGT 105
Ov 21 LeuAspAsnAsnGlyThrAlaThrProGlyLeuProThrGlnGlyVthrPheSerAsnVal 40

```
|||||
106 CTTGACAACAACGGTACTGTCACCCAGAGTTACTACCCAGGGAACATTTCAAAATGTG 165
QY 41 SerThr-AsnValSerTyrGlnGlu---ThrThrThrProSerThrLeuGlySerThrSe 59
Db 166 TCTACAAAGGTATCTCTACCAAGAAAGTACAAGACCTAGGTACCCTTGTAAAGTACCAG 225
QY 59 rLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVa 79
Db 226 CTGACCCCTGTCTCAACATGGCAATGAGGCCACAACAACATCACAGAACGACAGT 285
QY 79 lLysPheThr-SerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValG 99
Db 286 CAAATTCAGCATCTACTCTGTGATAACCTCAGTTTATGGAAACACAAACTCTTCTGTCC 345
QY 99 lnsrGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProG 119
Db 346 AGTCACAGACCTCTGTAATCAGACACAGTGTTCACCCAGCCCAACGTTTCAACTCCAG 405
QY 119 lnrThrThr-LeuLysProSerLeuSerProGlyAsnValSerAsp-LeuSerThrThr-S 138
Db 406 AGACAGCCTTGAAGCTAGCCTGTACCTGGAAATGTTTCAGACCCCTTTCACACCTAG 465
QY 138 erThrSerLeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerA 158
Db 466 GCACCTAGCCTTGCACATCTCCCACTAAACCCCTATACATCATCTTCTCTATCTTAAGTG 525
QY 158 spileLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleC 178
Db 526 ACATCAAGGGCAGAAATCAAAATGTTCAAGGATCAGAGAAGTGAATTTGACTCAGGGCATCT 585
QY 178 ysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyL 198
Db 586 GCCTGAGCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAGGACAGGGGAGGGAC 645
QY 198 euAlaArgValLeuCysGlyGluGluGln-AlaAspAlaasp-AlaGlyAlaGlnVal-- 216
Db 646 TGCCCGAGTGTGTGGGGAGGAGCAGAGCTGATGCTGATAGCTGGGGGCCAGGTATA 705
QY 217 -CysSerLeuLeuAlaGlnSer-GluValArgProGlnCysLeu-LeuValLeu 235
Db 706 GCTCCCTGCTACCTTGGCCAGCTCTAGAGGTGAGGCTCAGTGTCTACTGACTGGTCCCTG 765
QY 236 -AlaAsnArgThrGluIleSerSerLys 244
Db 766 GGCAACAGAAACAGACATTTACGAAAA 793
RESULT 10
AUI35182 754 bp mRNA linear EST 24-OCT-2000
LOCUS AUI35182 PLACE1 Homo sapiens cDNA clone PLACE1001404 5', mRNA
DEFINITION AUI35182 sequence.
ACCESSION AUI35182
VERSION AUI35182.1 GI:10995721
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 754)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
```

```
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES             Location/Qualifiers
     source           1..754
     organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="PLACE1001404"
     /clone_lib="PLACE1"
     /tissue_type="placenta"
     /note="Vector: pME18SFL3"
BASE COUNT      201 a      210 c      172 g      166 t      5 others
ORIGIN
Alignment Scores:      2.46e-75      Length:      754
Pred. No.:            991.50          Matches:    208
Score:                90.60%          Conservative: 4
Percent Similarity:   88.89%          Mismatches:  12
Best Local Similarity: 52.32%          Indels:     10
Query Match:          9               Gaps:       3
DB:
US-09-836-602-2 (1-373) x AUI35182 (1-754)
QY 1 MetProArgGlyThrPheAlaLeuCysLeuLeuSerLeuProSerGlyPheMetSer 20
Db 48 ATGCCCGGGGCTGGACCGCGCTTTGCTGCTGAGTTTGTGCTGCTTCTGGTTTCATGAGT 107
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 108 CTTGACAACAACGGTACTGCTACCCAGAGTTTACCTACCCAGGGAACATTTTCAAAATGTT 167
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 168 TCTACAAATGTTATCTCCCAAGAAACTACACACCTAGTACCTTGGAGTACACGCTG 227
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 228 CACCCCTGTCTCAACATGGCAATGAGGCCACCAACAACATCACAGAAAGGACAGTCAA 287
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 288 TTCACATCTACCTTCTGTGATAACCTCAGTTTATGGAAACACAAACTCTTCTGTCCAGTCA 347
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 348 CAGACCTCTGTANTCAGCAGAGTGTTCACACCCCGGCAACGTTTCAACTCCAGAGACA 407
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 408 ACCTTGAAGCCTAGCCTGTACACCTGGAAATGTTTCAGACCTTTCAACCCACTAGCCTAGC 467
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 468 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCCTATCTTAAGTGTACATCAAG 527
QY 161 Ala-GluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuG1 180
Db 528 GCAAGAAATCAAAATGTTTACGGCATCAGAGAAAGTGAATTCACACGGGCGCATCTGCCCTGGA 587
QY 180 uGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu---GlyLeuAl 199
Db 588 GCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAGGCAAGGGGAAAGGGCCCTGGGC 647
QY 199 aArgVal---LeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSe 218
Db 648 CCGAGTTGCTGTGTGGGAGGAGGAGGAGGCTGATGCC-----TG 686
QY 218 rLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeu 231
Db 687 ATTGCTGGGGGCCCAAGTTTANNNGCTCCCTGCTGCTCCCTT 726
RESULT 11
```

BG256232 602369196F1 NIH_MGC_91 1001 bp mRNA linear EST 13-FEB-2001
 LOCUS 602369196F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4477318 5',
 mRNA sequence.
 ACCESSION BG256232
 VERSION BG256232.1 GI:12766048
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1. (bases 1 to 1001)
 AUTHORS NIH-GSC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-rc@mail.nih.gov
 Tissue Procurement: DCTD/DTP
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10305 row: p column: 23
 High quality sequence start: 10
 High quality sequence stop: 579.

FEATURES

source Location/Qualifiers
 1..1001
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4477318"
 /clone_lib="NIH_MGC_91"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage resistant)"
 /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.4 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 283 a 287 c 257 g 174 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6.16e-68 Length: 1001
 Score: 907.50 Matches: 209
 Percent Similarity: 89.87% Conservative: 4
 Best Local Similarity: 88.19% Mismatches: 15
 Query Match: 47.89% Indels: 10
 DB: 10 Gaps: 1

US-09-836-602-2 (1-373) x BG256232 (1-1001)

Qy 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 72 ATGCGGGGGCTGGACCGCTTTGGTTGCTGAGTTGCTGCCTCTCGGTTTCATGAGT 131
 Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 132 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAATGTT 191
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 192 TCTACAATGATCTCTACCAAGAACTACACACCTAGTACCTTGGAGTACCAGCGCTG 251
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVallys 80
 Db 252 CACCCCTGTCTCAACATGGCAATGGAGCCCAACAACATCACAGAACACAGTCAAA 311
 Qy 81 PheThrSerThrValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 312 TTCACATCTACCTCTGTGATAACCTCACTTTATGGAAACACAAAACCTCTGTGTCAGTCA 371

Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 Db 372 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCAACGTTTCAACTCCAGAGACC 431
 Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThr-ThrSerThrSe 140
 Db 432 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTACAGACCTTTCACCAACATGACACTAG 491
 Qy 140 rLeuAla-ThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIle- 159
 Db 492 CTTTGCAAAACATCCCACTAAACCTATACATCATCTTCTCCTATCTCTAGTGACATCA 551
 Qy 160 LysAlaGluIle-LysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLe 179
 Db 552 AAGCCGCAAAATCAAAATGTTTCAGGCATCAGAGAAAGTGAACCTTGACTCA-GGCATCTGACT 610
 Qy 179 uGluGlnAsnLysThrSerSer-Cys-AlaGluPheLysLys--AspArgGlyGluGlyL 198
 Db 611 GGAGCACAAATAACACCTCCAGCTTGTTCGGAGGTTTAAAGAAAGGACAGAGGAGGGGCG 670
 Qy 198 euAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysS 218
 Db 671 TGGCCCGACGAGCTGTGTGGGAGGAGGAGGCTGATGCCAATGCGGGGCCCCAGGATGCT 730
 Qy 218 erLeuLeuAlaGlnSerGluValArgProGln 229
 Db 731 CCCCGCTCATGACC---AGTCGAGGAGGAGGCCAGG 762

RESULT 12

AU121843 717 bp mRNA linear EST 19-OCT-2000
 LOCUS AU121843 MAMMAL Homo sapiens cDNA clone MAMMAL1001088 5', mRNA
 DEFINITION sequence.

ACCESSION AU121843
 VERSION AU121843.1 GI:10937078
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1. (bases 1 to 717)
 AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
 Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
 Isogai, T.

TITLE

HRI human cDNA project
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp

FEATURES

source Location/Qualifiers
 1..717
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MAMMAL1001088"
 /clone_lib="MAMMAL"
 /tissue_type="mammary gland"
 /note="Vector: pWE185FL3"

BASE COUNT 195 a 208 c 164 g 147 t
 ORIGIN

Alignment Scores:

Pred. No.: 3.23e-58 Length: 717
 Score: 792.00 Matches: 163
 Percent Similarity: 97.02% Conservative: 0
 Best Local Similarity: 97.02% Mismatches: 5

Query Match: 41.79% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AU121843 (1-717)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
DB 214 ATGCCGGGGCTGGAGCGGCTTGGCTGCTGAGTTGGCTTCTGGGTTTCATGAGT 273

QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 274 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 333

QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 334 TCTACAATGATCTCCACCAAGAACTACACACCTAGTACCCCTTGGAGTACACGCTG 393

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
DB 394 CACCCCTGTGCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA 453

QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 454 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAACTCTTCTGTCCAGTCA 513

QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 514 CAGACCTCTGTAATCAGCAGAGTTTCACCCACCCAGCCNACGTTTCACTCCAGAGACA 573

QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
DB 574 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTCAGACCTTTCACACCTAGCCTAGC 633

QY 141 LeuAlaThrSerProThrLysProTyrThrThrSerSerSerProIleLeuSerAspIleLys 160
DB 634 CTTGCAACATCTTCCACTAACCCCTATCA-TCATCTTCTNCTATNCTAAGTGACATCAAG 692

QY 161 AlaGluIleLysCysSerGlyIle 168
DB 693 GCAGAAATCAATGTCANGCATC 716

RESULT 13
AL547847
LOCUS AL547847 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI033YP01 5
DEFINITION prime, mRNA sequence.
ACCESSION AL547847
VERSION AL547847.1 GI:12882292
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 997)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .997
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI033YP01"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-Oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com

BASE COUNT 236 a 289 c 248 g 221 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 6,34e-58 Length: 997
Score: 791.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.74% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AL547847 (1-997)

QY 218 SerLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsn 237
DB 2 TCCTCTGCTCTTCCGCCAGTCTGAGGTGAGGCCTCAGTGTCTACTGTGTTGGCCAAC 61

QY 238 ArgThrGluIleSerSerLysLeuGlnLeuMetLysHisGlnSerAspLeuLys 257
DB 62 AGAACAGAAATTTCCAGCAACTCCAACTTATGAAAAGCACCACCAATCTGACCTGAAAG 121

QY 258 LeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLys 277
DB 122 CTGGGATCTCTAGATTCTAGTGAAGATGTTGCAAGCCACGAGCTATTTCCTCCAAAG 181

QY 278 ThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyr 297
DB 182 ACCCTGATTGCTACTGCTCAGCTCGGAGCCCTGCTGCTGCTTGGCATCACTGGCTAT 241

QY 298 PheLeuMetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyr 317
DB 242 TTCTCTGATGAATCGCGCAGCTGGAGCCCCACAGAGAAAGGCTGGCGAAGACCCCTAT 301

QY 318 TyrThrGluAsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAla 337
DB 302 TACACGGAACCGGTGGAGCCAGGCTATAGCTCAGGACCTGGGACCTCCCTGAGGCT 361

QY 338 GlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSer 357
DB 362 CAGGGAAGGCCAGTGTGAACCGAGGGCTCAGAAAACGGGACCGCCAGGCCACCTCC 421

QY 358 ArgAsnGlyHisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
DB 422 AGAACGGGCATTTCAGCAAGACAAACACGCTGGTGGCTGATACCGAATTG 469

RESULT 14

AU141125

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 733)

1 (bases 1 to 733)

1 (bases 1 to 733)

1 (bases 1 to 733)

1 (bases 1 to 733)

1 (bases 1 to 733)

1 (bases 1 to 733)

1 (bases 1 to 733)

1 (bases 1 to 733)

1 (bases 1 to 733)

1 (bases 1 to 733)

Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES
 source
 1. .733
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="THYR01000063"
 /clone_lib="THYR01"
 /tissue_type="thyroid gland"
 /note="Vector: pME18SFL3"

BASE COUNT 199 a 212 c 168 g 150 t 4 others
 ORIGIN

Alignment Scores:
 Pred. No.: 4,59e-55 Length: 733
 Score: 755.50 Matches: 157
 Percent Similarity: 94.61% Conservative: 1
 Best Local Similarity: 94.01% Mismatches: 8
 Query Match: 39.87% Indels: 2
 DB: 9 Gaps: 1

US-09-836-602-2 (1-373) x AU141125 (1-733)

Qy 1 MetProArgGlyTTPThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 233 ATGCCGGGGCTGGACCGCGCTTTGCTGTGAGTTGCTGCTTCTGGGTTTCATGAGT 292
 Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 293 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAAATGTT 352
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 353 TCTACAAATGTATCTACCAAGAAACTACACACCTAGTACCTTGGAAAGTACCAGCGCTG 412
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
 Db 413 CACCCTGTGCTCAACATGGCAATGAGCCCAACAACATCACAGAAACGACAGTCAAA 472
 Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 473 TTCACATCTACCTCTGTGATACCTCAGTTTATGGAACACAAACTCTCTGTCCAGTCA 532
 Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 Db 533 CAGACCTCTGTAATCAGCACAGTGTTCACCACCCAGCCAACTTTCAACTCCAGAGACA 592
 Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 Db 593 ACCTTGAAGCCCTANCCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTACACTACCT 652
 Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
 Db 653 TGCAACATCTCC---ACTAAACCTCTATCA-TCATCTTCTNCTATCTTAAAGTGACATCAAG 708
 Qy 161 AlaGluLeuLysCysSerGly 167
 Db 709 GCNGAAATCAAATGTTTCAGGC 729

RESULT 15
 LOCUS AU139201 825 bp mRNA linear EST 25-OCT-2000
 DEFINITION AU139201 PLACE1 Homo sapiens cDNA clone PLACE1010139 5', mRNA
 sequence.
 ACCESSION AU139201
 VERSION AU139201.1 GI:11000722
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 825)
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isogai,T.

TITLE
 JOURNAL
 COMMENT
 HRI human cDNA project
 Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES
 source
 1. .825
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="PLACE1010139"
 /clone_lib="PLACE1"
 /tissue_type="placenta"
 /note="Vector: pME18SFL3"

BASE COUNT 227 a 243 c 185 g 165 t 5 others
 ORIGIN

Alignment Scores:
 Pred. No.: 2.01e-51 Length: 825
 Score: 714.00 Matches: 148
 Percent Similarity: 95.54% Conservative: 2
 Best Local Similarity: 94.27% Mismatches: 4
 Query Match: 37.68% Indels: 3
 DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AU139201 (1-825)

Qy 1 MetProArgGlyTTPThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 301 ATGCCGGGGCTGGACCGCGCTTTGCTGTGAGTTTGTGCTTCTGGGTTTCATGAGT 360
 Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 361 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAAATGTT 420
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 421 TCTACAAATGTATCTACCAAGAAACTACAAACCTAGTACCTTGGAAAGTACCAGCGCTG 480
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
 Db 481 CACCCTGTGCTCAACATGGCAATGAGCCCAACAACATCACAGAAACGACAGTCAAA 540
 Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 541 TTCACATCTACCTCTGTGATACCTCAGTTTATGGAACAAACATCTCTGTCCAGTCA 600
 Qy 101 Gln-ThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 Db 601 CAAGACCTCTCTAATCAGCACAGTGTTCANCAACCCACCAACCTTTCACTCCAGAGAC 660
 Qy 120 rThrLeuLysProSerLeuSerProGly-AsnValSerAspLeuSerThrThrSerThrS 140
 Db 661 AACCTTGAAGCCCTAACCTGTACCTGGAAATGTTTCAAACCTTTTAACTACCACTANCACTA 720
 Qy 140 rLeuAlaThrSerProThr-LysProTyrThrThrSerSerPro 154
 Db 721 GCCTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTCT 765

Search completed: October 30, 2002, 10:47:52
Job time : 1301.57 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 30, 2002, 08:08:13 : Search time 47.8495 Seconds
(without alignments)
1914.781 Million cell updates/sec

Title: US-09-836-602-2
Perfect score: 1895
Sequence: 1 MPRGWTALCLLSLPSGFMS.....QATSRNGHSARQHVADTEL 373

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09836602/runat_29102002_101145_3837/app_query.fasta_1.1429
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602 -CGN1_1_44 -runat_29102002_101145_3837 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6C.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	904	47.7	1065	2	US-08-475-634D-18
2	144	7.6	5163	3	US-08-700-651-1
3	144	7.6	5163	3	US-08-928-361B-4
4	144	7.6	5318	3	US-08-700-651-2
5	144	7.6	5318	3	US-08-928-361B-3
6	138	7.3	5511	3	US-08-928-361B-2
7	138	7.3	7334	3	US-08-928-361B-1
8	129	6.8	1505	1	US-07-915-246-1
9	128	6.8	390	4	US-09-197-649-7
10	127.5	6.7	3337	1	US-08-072-610-1
11	127.5	6.7	3337	2	US-08-719-822B-1
12	127.5	6.7	3337	4	US-09-092-458-1

13	125.5	6.6	2589	1	US-08-325-267A-3	Sequence 3, Appli
14	125.5	6.6	2685	3	US-08-362-525-21	Sequence 21, Appl
15	125.5	6.6	4614	1	US-08-325-267A-1	Sequence 1, Appli
16	123	6.5	2754	1	US-08-270-076A-10	Sequence 10, Appl
17	121	6.4	2214	6	5258502-1	Patent No. 5258502
18	121	6.4	3168	4	US-09-165-239A-3	Sequence 3, Appli
19	119.5	6.3	688	4	US-08-998-416-915	Sequence 915, App
20	119.5	6.3	1107	2	US-08-991-300-1	Sequence 1, Appli
21	119.5	6.3	80161	3	US-09-036-987A-1	Sequence 1, Appli
22	119.5	6.3	80161	4	US-09-370-700-1	Sequence 1, Appli
23	118	6.2	2150	2	US-08-861-464-13	Sequence 13, Appl
24	118	6.2	2150	4	US-08-396-001-13	Sequence 13, Appl
25	118	6.2	2150	4	US-09-323-433A-13	Sequence 13, Appl
26	117	6.2	2584	3	US-08-758-662-8	Sequence 8, Appli
27	116	6.1	2793	1	US-08-209-747-1	Sequence 1, Appli
28	116	6.1	2793	1	US-08-458-298-1	Sequence 1, Appli
29	115.5	6.1	2032	4	US-09-241-581B-5	Sequence 5, Appli
30	115.5	6.1	2032	5	PCT-US95-07721-5	Sequence 5, Appli
31	114.5	6.0	2093	1	US-08-287-001A-1	Sequence 1, Appli
32	114.5	6.0	2093	5	PCT-US95-09941-1	Sequence 1, Appli
33	113.5	6.0	2164	4	US-08-760-615-3	Sequence 3, Appli
34	113	6.0	36519	3	US-08-923-137-2	Sequence 2, Appli
35	112.5	5.9	4031	2	US-08-993-118-1	Sequence 1, Appli
36	112.5	5.9	4031	3	US-08-845-528C-1	Sequence 1, Appli
37	112.5	5.9	4225	2	US-08-993-118-9	Sequence 9, Appli
38	112.5	5.9	4225	3	US-08-845-528C-9	Sequence 9, Appli
39	112.5	5.9	4265	4	US-09-061-709-1	Sequence 1, Appli
40	111	5.9	3181	1	US-08-655-086-1	Sequence 1, Appli
41	110	5.8	6027	2	US-08-968-542C-1	Sequence 1, Appli
42	109.5	5.8	3833	1	US-08-917-320-18	Sequence 18, Appl
43	109.5	5.8	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
44	109.5	5.8	5931	3	US-08-783-774-1	Sequence 1, Appli
45	108.5	5.7	4108	4	US-08-981-729-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-475-634D-18
; Sequence 18, Application US/08475634D
; Patent No. 5962644
; GENERAL INFORMATION:
; APPLICANT: HAWLEY, Robert J.
; APPLICANT: MONROY, Rodney L.
; TITLE OF INVENTION: Antibodies to Porcine CD34 Positive
; TITLE OF INVENTION: and Uses Therefor
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS 6.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475.634D
; FILING DATE: June 7, 1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24, 025
; REFERENCE/DOCKET NUMBER: 61750-141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 973-994-1700
; TELEFAX: 973-994-1744
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 1065 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-475-634D-18

Alignment Scores:
Pred. No.: 1,06e-84 Length: 1065
Score: 904.00 Matches: 196
Percent Similarity: 69.44% Conservatives: 38
Best Local Similarity: 58.16% Mismatches: 77
Query Match: 47.70% Indels: 26
DB: 2 Gaps: 6

US-09-836-602-2 (1-373) x US-08-475-634D-18 (1-1065)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 67 ATGCCCGGGGTGGACCGCTCTGCTGCTGAGTTGCTGCCCTCTGGGTTTCACAGCT 136
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuPro----- 32
Db 127 GTG---AACAGCTCAACTATTGCTTCCACCTTGGCCAGCTGCCGCTGGTCAACTCCACC 183
QY 33 -----ThrGlnGlyThrPheSerAsn----- 39
Db 184 GGGCCGGTACCCGAGGGCAGTATACCGGGTCAACTATCTCACATATCTTCACTT 243
QY 40 ValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSer 59
Db 244 GTTCTTACAAATATATCAACGAGGAAACACATCA---GATGCTTTCGAAAGTGCAGC 300
QY 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVal 79
Db 301 CTCACACATGCTCTCAGGGCAGCAGTGGGACACCGTAGCCATCTCAGGCCCTACAGT 360
QY 80 LysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGln 99
Db 361 AATTTTCATGCTACCTCGGGGTCACTCGTCCCGGAAACCGTTAACTTCTGTCCAG 420
QY 100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
Db 421 CCTCAGACCTCT---CTAGCCACAGCGTCTCTCGGCCACCATCACTTACAACTTCAGAG 477
QY 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 139
Db 478 GTGACCTCGACCCCGACAGCTTCCAGGAAATGTTTCAGACCCCTCTACACAGTACC 537
QY 140 SerLeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIle 159
Db 538 AGCCCTGCGAGATCCCCACCGCTTACACATCATCTCTCTCTACCCCGGATGCCAC 597
QY 160 LysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Db 598 AAGGGGAAGTCAAAATGTGCCAAATCAAGAGGTCAAAATGACCAAGGTATCTGCTG 657
QY 180 GluAlaAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Db 658 GACGGAATGAGACCTCCGGCTCGGAGAGTTTAAAGAGGACAATGGAGAGAGTTGATG 717
QY 200 ArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
Db 718 CAAATCCTGTGGGCGAGGAGGCTGAGCGGCGCCAGGG-----GTGTGCTCTTG 771
QY 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThr 239
Db 772 CTCCTTGCCTCAATCTGAGGTGAACCTCACTGCTGCTGCTGTGGCCACGGAACA 831
QY 240 GluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGly 259
Db 832 GAATTAGCAGCANGTCTCTCTGGGAAAGCACCAGTCTGAAGTCTGAGAGAGTACG 891
```

```
QY 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu 279
Db 892 ATCAAAACTTCTCGAAACAAGATGTTAGGAGCCACAGAGCTACTCCCGAAAGACCTTG 951
QY 280 IleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeu 299
Db 952 ATTGCACCTGGTCACTACCTCGGGATCCTGCTGGCTGTCTTGGCATCACTGGCTACTT 1011
QY 300 MetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspPro 316
Db 1012 ATGAACCGTCGCACTGGAGCCCTACAGGAGAAAGGCTGGAGCTGGAACCC 1062
RESULT 2
US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4 (HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5163
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-1

Alignment Scores:
Pred. No.: 0.000241 Length: 5163
Score: 144.00 Matches: 37
Percent Similarity: 47.37% Conservatives: 26
Best Local Similarity: 27.82% Mismatches: 69
Query Match: 7.60% Indels: 1
DB: 3 Gaps: 0

US-09-836-602-2 (1-373) x US-08-700-651-1 (1-5163)

QY 22 AspAsnAsnGly-ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 41
Db 613 GACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 672
QY 41 rThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeuH 61
Db 673 GACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 732
QY 61 sProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLysPh 81
Db 733 AAGCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 792
QY 81 eThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGl 101
Db 793 TACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 852
QY 101 nThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrTh 121
Db 853 AACCAACTACTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 912
QY 121 rLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrThrSerThrSerLe 141
Db 913 CAAGAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 972
QY 141 uAlaThrSerProThrLysProTyrThrSerSerSer 153
```



```
Qy 184 ThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArgValLeuLysCys 203
Db 3295 ---GGCGAATGT-----GAGCAGAAAGGAGCAACTTATGTTGGTGT 3333
Qy 204 GlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGln 223
Db 3334 ATCGGAAAGATGACGATATTGAAATGGAATGTCATTTACAATGATTTCCAAATGATGAC 3393
Qy 224 SerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGluIle----- 241
Db 3394 ACGCATGTCGTTTCAGATTTAAGTAAAGATGATGAGACACTATTTCAGTAAAGATGC 3453
Qy 242 -----SerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeu 258
Db 3454 GGAAGAGTGCAGGTAACCTGAGTTCCAGATAGAAGT----- 3492
Qy 259 GlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTySerGlnLysThr 278
Db 3493 -----TTGGATTTTCACAATTTCCCTCCAGTAGCTGGCCATACACGCTGTTCAC----- 3537
Qy 279 LeuIleAlaLeuValThrSerGlyAlaLeuAlaValLeuGlyIleThrGlyTyrPhe 298
Db 3538 -----ATATAGTTGGTGTGAGCGGC----- 3558
Qy 299 LeuMetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyr 317
Db 3559 -----GGTGGAATAATTCAGTAAGCCCATAC 3585

RESULT 8
US-07-915-246-1/c
; Sequence 1, Application US/07915246
; Patent No. 5401836
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Chris L.
; APPLICANT: Fallis, Lynne
; APPLICANT: Bellmare, Guy
; APPLICANT: Boivin, Rodolphe
; TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
; TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held, and Malloy
; STREET: 500 W. Madison St. 34th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/915,246
; FILING DATE: 19920716
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 91 P 1125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 707-8889
; TELEFAX: 312 707-9155
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1505 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
```

```
; ORGANISM: Brassica napus
; STRAIN: Westar
; DEVELOPMENTAL STAGE: Somatic
; TISSUE TYPE: Root
; US-07-915-246-1

Alignment Scores:
Pred. No.: 0.00104 Length: 1505
Score: 129.00 Matches: 47
Percent Similarity: 48.78% Conservatives: 33
Best Local Similarity: 28.66% Mismatches: 70
Query Match: 6.81% Indels: 14
DB: 1 Gaps: 5

US-09-836-602-2 (1-373) x US-07-915-246-1 (1-1505)

Qy 12 SerLeuLeuProSerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeu 31
Db 1370 AGCTGCATAACCTCCCTCCTCCTCCGCGCACCTCCAGCTCCACACCGCGCTCCTCC 1311
Qy 32 ProThrGlnGlyThrPheSer-----AsnValSerThrAsnValSerTyr 46
Db 1310 ACCACCTGTCTCTCTCTCCACACACACCGTATCCCTCTCCAGCTCCACCTCCGGC 1251
Qy 47 GlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSerGlnHis 66
Db 1250 ACCATAACCAACCGGTGAGCTCCACCGCGCGGCGGCTCTCTCTCCACCGCTCTTT 1191
Qy 67 GlyAsnGluAlaThrThrAsnIle-----ThrGluThrThrValLysPheThrSer 83
Db 1190 TCCCGCGCTCCACACCTCCATGTCCACCTGCACCGCGCACCTCCATATCTCTCCAGC 1131
Qy 84 ThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGlnThrSer 103
Db 1130 ACCAGCGCTTCTCCACACCGGTATCCACCGCTGTCACCTCCAGCTCTCTCCACCGCC 1071
Qy 104 ValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrThrLeuLys 123
Db 1070 TCCATTTCCACCGCTCCACCAAC-----TCCATGTCACCTGCACCGCGCAC----- 1023
Qy 124 ProSerLeuSerProGlyAsnValSerAspLeuSerThrThr-----SerThrSerLeuAla 142
Db 1022 ---TCCATATCTCTCTCCAGCACCGACCTTCTCCACCGCGGTATCCACCGCTCCAGC 966
Qy 143 ThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLysAlaGlu 162
Db 965 ACCGCGAGCTCTCCACCAACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 906
Qy 163 IleLysCysSer 166
Db 905 TCCCTCCGCTCC 894

RESULT 9
US-09-197-649-7
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Pribnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/G1-CON
; CURRENT APPLICATION NUMBER: US/09/197,649
; CURRENT FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: 07/829,461
; EARLIER FILING DATE: 1992-01-31
; EARLIER APPLICATION NUMBER: 07/739,055
; EARLIER FILING DATE: 1991-08-01
; EARLIER APPLICATION NUMBER: 07/561,968
; EARLIER FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
```



```
Db 1765 ACAGCCTCTC---CTGCATCAGGA--TCCTTGAGGATT-----TGTATAAATAATGGCTC 1716
Qy 265 uGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIleAlaLeuValThrSe 285
Db 1715 AAAGGATGTAATCTCGCATACGTTAAATCGTTATCTAAGATATCAACGAATAAAGTTTC 1656
Qy 285 rGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMetAsnArgArgSerTr 305
Db 1655 ATCTGCCGCTTGAGGACCATGTTTTCACAGGCTGACAGGAGTCACACTACTCTACCGG 1596
Qy 305 pSer---ProThrGlyGluArgLeuGlyGluAspProTyrTyrThrGluAsnGlyGlyG 324
Db 1595 TTCGCCCACTACTGGCACTGCTACTAAGCCTTCTCCT-----TC 1557
Qy 324 yGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLysAlaSerValAs 344
Db 1556 TCCCTCCCTCTCCTCTCTCTCTCCCTTCTCCCTTCTCCTGATGTTCTCTTAAT----- 1502
Qy 344 nArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGlyHisSerAla 363
Db 1501 -----TGGAAGTCATCCTCTGGAGTTCGCTTAACCTTCTCTCTCTCTGCA 1454
```

RESULT 12

```
US-09-092-458-1/c
; Sequence 1, Application US/09092458
; Patent No. 6231861
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,458
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/719,821
; FILING DATE: 09/30/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pVMB3.3.1
; US-09-092-458-1
```

```
Alignment Scores:
Pred. No.: 0.00592 Length: 3337
Score: 127.50 Matches: 93
Percent Similarity: 37.89% Conservative: 51
Best Local Similarity: 24.47% Mismatches: 194
Query Match: 6.73% Indels: 43
DB: 4 Gaps: 13

US-09-836-602-2 (1-373) x US-09-092-458-1 (1-3337)

Qy 5 TrpThrAlaLeuCysLeuLeuSerLeuLeuPro-----SerGlyPheMetSerLeu 21
Db 2527 TGGTAGTCATCGATCGATAGATAAAATCCAGAGCCATAATTCGTCTCTTTGTTGTTG 2468
Qy 22 AspAsn-----AsnGlyThrAlaThrProGluLeuProThrGlnGlyThr 36
Db 2467 TCTAATGATATCTCTCCGATGGTATTACTAATTGTAATACATCTCTTCCTCTACT 2408
Qy 37 PheSerAsnValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGly 56
Db 2407 GGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2348
Qy 57 SerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrAsnIleThrGlu 76
Db 2347 GGTACTTCTACTTCTACTACGCTGGTACTCTACTTCTCTACTTCTCTCCACTTCTCC 2288
Qy 77 ThrThrValLysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSer 96
Db 2287 ACTTCTCTGGTACTCTTCCACTTCTCTGGTACTCTCTCCACTCTCTCTGGTACTCTCT 2228
Qy 97 SerValGlnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSer 116
Db 2227 TCCACTTCTCTGGTACTCT-----TCCACTTCTCTACTCTCTCGGGTACTCTCT 2174
Qy 117 ThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThr 136
Db 2173 ACTTCTCTGGTACTCTTCCACTTCTCTACTTCTGAGGTACTCT-----TCCACT 2120
Qy 137 ThrSerThrSerLeuAlaThrSerProThr---LysProTyrThrSerSerSerProIle 155
Db 2119 TCTTCTACTTCTGAGGTACTCTTCCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2063
Qy 156 LeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGln 175
Db 2062 --TCTACGGTTCTCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2006
Qy 176 GlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGly 195
Db 2005 GGAGTTGCTCTAATCTCTTCTCTGGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1946
Qy 196 GluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGln 215
Db 1945 TCTTCTCTGGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1886
Qy 216 ValCysSer-LeuLeu-----LeuAlaGlnSerGluValArgProGlnCysLe 231
Db 1885 GCAGTTCTCTAATTCGAAATCGCTCTCTGGAGTTGCATCAATCTTCCGTTGGCCT 1826
Qy 231 uLeu-----LeuValLeuAlaAsnArgThrGluIleSerSerLysLe 245
Db 1825 TCTTCCGGGGGCCCTTACTGCCACTGGTACTTGTACAGGTGCTTCTCTCTCTCTCTCT 1766
Qy 245 uGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIleLeuAspPheThrG 265
Db 1765 ACAGCCTCTC---CTGCATCAGGA--TCCTTGAGGATT-----TGTATAAATAAGGCTC 1716
Qy 265 uGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIleAlaLeuValThrSe 285
Db 1715 AAAGGATGTAATGCTCTGCATACGTTAAATCGTTATCTAAGATATCAACGAATAAAGTTTC 1656
Qy 285 rGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMetAsnArgArgSerTr 305
Db 1715 AAAGGATGTAATGCTCTGCATACGTTAAATCGTTATCTAAGATATCAACGAATAAAGTTTC 1656
Qy 285 rGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMetAsnArgArgSerTr 305
Db 1715 AAAGGATGTAATGCTCTGCATACGTTAAATCGTTATCTAAGATATCAACGAATAAAGTTTC 1656
```



```
Qy 294 IlerThrGlyTyrPheLeuMetAsnArgArgSerTrpSerPro-----ThrGlyGluArg 311
Db 1963 GTTACTACAGATAC-----ACACATGGTGTCTCTATTTCACACAGAAATCG 2010
Qy 312 LeuGlyGluAspProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGly--- 330
Db 2011 AGGCAACAACAAACGCTAGTTACTTCTCTCGAATCTGGTGTGTCCGAAACT 2070
Qy 330 ----- 330
Db 2071 GCTTCACCTGCCATGTTTCGAGGCCACGCTACTGTGAATGATGTTGTACGGTCTAT 2130
Qy 331 ProGlyThrSerProGluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsn 350
Db 2131 CCTACATGGAGGCCACAGACTCGGAATGAAGAGTCTGTGCAGC---TCTAAATGAACAGT 2187
Qy 351 GlyThrGlyGlnAlaThrSerArgAsn-----GlyHisSerAlaArgGlnHisValVal 368
Db 2188 GCTACCGGTGAGACAAACCACTTCTAGCTGTGCTGAACGACTACCAATACTGTAGCT 2247
Qy 369 AlaAspThr 371
Db 2248 GCTGAGACG 2256

RESULT 14
US-08-362-525-21
; Sequence 21, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLIS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 213285/T7020(V)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELE: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 2685 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: pY105
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2685
; OTHER INFORMATION: /product= "Flocculation protein" /gene= nFLOI"
US-08-362-525-21

Alignment Scores:
Pred. No.: 0.00655 Length: 2685
Score: 125.50 Matches: 100
Percent Similarity: 38.30% Conservative: 62
Best Local Similarity: 23.64% Mismatches: 148
Query Match: 6.62% Indels: 115
DB: 3 Gaps: 20

US-09-836-602-2 (1-373) x US-08-362-525-21 (1-2685)

Qy 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
Db 1234 ACAACCACCACTGAA---CCATGGAGCTGTACTTTTACTTCGACTTCACATCAAAATGCT 1290
Qy 46 -----TyrGlnGluThrThr-----ThrPro 52
Db 1291 ACTGTCACTGGAACCAATGGCTTGCCAACTGATGAACACTGTCATTGTTGTCAAAATCCA 1350
Qy 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
Db 1351 ACTACTGCCATCTCATCCAGTTGTCA-----TCATCATCTTCAGGACAAATCACCAGC 1404
Qy 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValThrGly 92
Db 1405 TCTATCAGC-----TCTTCGCGTCCAAATTTATCCCATCTCTATCTCT 1446
Qy 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
Db 1447 AGCAATGAACACTCTGTGATTCT---TCCTCAGTAATTTCT----- 1485
Qy 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
Db 1486 TCCTCAGTCACTTCTCTTATTCATCTTCTTCCAGTCATTTCTCTCAGTCATCTTCT 1545
Qy 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
Db 1546 TCTTCTACAACAACTCCACTCTATATTTCTGAATCATCTAAATCATCCGTCATTCCA 1605
Qy 146 ThrLysProTyrThrSerSerSerProIleLeuSerAspIleLysAlaGluLysCys 165
Db 1606 ACCAGTAGTTCACCTCTGGTTCT-----TCTGAGAGCGAAACG 1644
Qy 166 SerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSer 185
Db 1645 AGTTCAGCTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1704
Qy 186 -----SerCysAlaGluPheLysLysAspArgGly 195
Db 1705 ACATATTTCTTCTCATCATTTACCATTTGTACCGGACCAACAGCGAGAAACTGCT 1764
Qy 196 GluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAla----- 210
Db 1765 TCTTCTATAC---CACCTGCTACCACTACAAAAACGAGCGAACACCACTTTGGTTACCG 1822
Qy 211 ---AspAlaGlyAlaGlnValCysSerLeuLeuLeuLeuGlnSerGluValArgProGln 229
Db 1823 TGACATCCTCGAGTCTCATGTGTGCAGTGAATTCATCTCCCTCGGATTTGTTCCACAG 1882
```

```
QY 230 CysLeuLeuValLeuAlaAsnArgThrGluile---SerSerLysLeuGlnLeuMet 248
|||||
Db 1883 CTACTGTTACTGTTAGCGGCTCACACAGAGTATACCACATGTCCTTCTACTA 1942
|||||
QY 249 LysLysHisGlnSerAspLeuLysLeuGlyLeuAspPheThr----- 264
|||||
Db 1943 CAGAGACAACAAGC-----AAACCAAA-GGGACAACAGCAACACAGAAACAACA 1995
|||||
QY 265 -----GluGlnAspValAlaSerHisGlnSer 273
|||||
Db 1996 AAACAAACCCAGGAGTGAATTTCTTGTGAATCTGACGTATGCTCT----- 2046
|||||
QY 274 TyrSerGlnLysThrLeuileAlaLeuValThrSerGlyAlaLeuAlaValLeuGly 293
|||||
Db 2047 -----AAGACTGCTTCCAGCCATTGTATCTACAGCACTGCTACTATTAAACGGC 2097
|||||
QY 294 IleThrGlyTyrPheLeuMetAsnArgSerTrpSerPro-----ThrGlyGluArg 311
|||||
Db 2098 GTTACTACAGAAATAC-----ACAACATGGTGTCTCTATTTCACACAGAAATCG 2145
|||||
QY 312 LeuGlyGluAspProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGly--- 330
|||||
Db 2146 AGGCACAACAACAGCGTAGTTACTGTTACTCTCTGCGAAATCTGGTGTTCGAAACT 2205
|||||
QY 330 ----- 330
|||||
Db 2206 GCTTCACCTGCCATTGTTTCGACGCCACGGCTACTGTGAATGATGTTGTACGGTCTAT 2265
|||||
QY 331 ProGlyThrSerProGluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsn 350
|||||
Db 2266 CCTACATGGAGGCCACACAGACTGCGAATGAAGAGTCTCTCAGC---TCTAAATGAACAGT 2322
|||||
QY 351 GlyThrGlyGlnAlaThrSerArgAsn-----GlyHisSerAlaArgGlnHisValVal 368
|||||
Db 2323 GCTACCCGGTGAAGACACCAACCAATACTTTAGCTGCTGAAACGACTACCAATACTGTAGCT 2382
|||||
QY 369 AlaAspThr 371
|||||
Db 2383 GCTGAGACG 2391
|||||
RESULT 15
US-08-325-267A-1
; Sequence 1, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; NUMBER OF INVENTION: CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: ABXL-1D
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4614
; US-08-325-267A-1
Alignment Scores:
Pred. No.: 0.0168 Length: 4614
Score: 125.50 Matches: 100
Percent Similarity: 38.30% Conservative: 62
Best Local Similarity: 23.64% Mismatches: 148
Query Match: 6.62% Indels: 115
DB: 1 Gaps: 20
US-09-836-602-2 (1-373) x US-08-325-267A-1 (1-4614)
QY 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
|||||
Db 3124 ACAACCAACCACTGAA---CCATGAGCTGGTACTTTTACTTCGACTTCCACTGAAATGCT 3180
|||||
QY 46 -----TyrGlnGluThrThr-----ThrPro 52
|||||
Db 3181 ACTGTCTACTGGAACCAATGCTTGCACCACTGATGAAACTGTCATTTGTGTCAAACCTCCA 3240
|||||
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
|||||
Db 3241 ACTACTGCCATCTCATCCAGCTTTGTCA-----TCATCATCTTCAGGACAAATCACCAGC 3294
|||||
QY 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValTyrGly 92
|||||
Db 3295 TCTATCAGC-----TCTTCGCGTCCAATTATTACCCCATTTCTATCCT 3336
|||||
QY 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
|||||
Db 3337 AGCAATGGAACTCTCTGTGATTCT---TCCTCAGTAATTTCT----- 3375
|||||
QY 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
|||||
Db 3376 TCTCAGTCACTCTCTCTCTATTCACACTCTCTCCAGTCAATTTCTCTCCTCAGTCATTTCT 3435
|||||
QY 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
|||||
Db 3436 TCTTCTACAACAACCTCCACTCTCTATATTTTCTGAATCATCTAAATCATCCGTCATTTCCA 3495
|||||
QY 146 ThrLysProTyrThrSerSerSerProIleLeuSerAspIleLysAlaGluIleLysCys 165
|||||
Db 3496 ACCAGTAGTTCCACCTCTCTGTTCT-----TCTGAGACGGAACG 3534
|||||
QY 166 SerGlyIleArgGluGluValLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSer 185
|||||
Db 3535 AGTTCAGCTGGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3594
|||||
QY 186 -----SerCysAlaGluPheLysLysAspArgGly 195
|||||
```

```
Db 3595 ACATATTCTTCTTCATCATTACCACTTGTACCACTGCGACAACAAGCCAGGAACACTGCT 3654
      : : : : : ||| : : : : :
Qy 196 GluGlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAla----- 210
      ||| : : : : : ||| : : : : :
Db 3655 TCTTCATTAC--CACCTCTACCACTACAAAACGACGAAACAACCACTTTGGTTACCG 3712
      : : : : : : : : : : : : : : :
Qy 211 ---AspAlaGlyAlaGlnValCysSerLeuLeuAlaGlnSerGluValArgProGln 229
      : : : : : ||| : : : : :
Db 3713 TGACATCCTGCGAGTCTCTGTCATCTGAATCCATCTCCCTCGGATTGTTCCACAG 3772
      : : : : : ||| : : : : :
Qy 230 CysLeuLeuValLeuAlaAsnArgThrGluIle---SerSerLysLysGlnLeuMet 248
      ||| : : : : : ||| : : : : :
Db 3773 CTACTGTTACTGTTAGCGCGCTCACACAGAGTATACCACATGGTGCCTATTCTACTA 3832
      : : : : : : : : : : : : : : :
Qy 249 LysLysHisGlnSerAspLeuLysLysLeuGlyIleLeuAspPheThr----- 264
      : : : : : ||| : : : : :
Db 3833 CAGAGACAACAANGC-----AAACCAAA-GGGACAACAGAGCAACCCACAGAACACA 3885
      : : : : : : : : : : : : : : :
Qy 265 -----GluGlnAspValAlaSerHisGlnSer 273
      : : : : : ||| : : : : :
Db 3886 AAACAACACCGGTAGTTACAATTTCTTCTGTGAATCTGACGTATGCTCT- 3936
      : : : : : ||| : : : : :
Qy 274 TyrSerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGly 293
      ||| : : : : : ||| : : : : :
Db 3937 -----AAGACTGCTTCTCCAGCCATTGTATCTACAAGCAGCTGCTACTATTAAACGGC 3987
      : : : : : : : : : : : : : : :
Qy 294 IleThrGlyTyrPheLeuMetAsnArgSerTrpSerPro-----ThrGlyGluArg 311
      : : : : : : : : : : : : : : :
Db 3988 GTTACTACAGAATAC-----ACAACATGGTGTCTCTATTTCACCACACAGAAATCG 4035
      : : : : : ||| : : : : :
Qy 312 LeuGlyGluAspProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGly--- 330
      : : : : : ||| : : : : :
Db 4036 AGGCACAACAACACGCTAGTTACTGTTACTTCCCGAATCTGGTGTGTTCGAAACT 4095
      : : : : : ||| : : : : :
Qy 330 ----- 330
      : : : : : : : : : : : : : : :
Db 4096 GCTTCACCTGCCATTGTTTCGAGGCCACGCTACTGTGAATGATGTTGTACGGTCTAT 4155
      : : : : : : : : : : : : : : :
Qy 331 ProGlyThrSerProGluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsn 350
      ||| : : : : : ||| : : : : :
Db 4156 CCTACATGGAGGCCACAGACTCGGAATGAGAGTCTGTCTAGC---TCTAAAATGAACAGT 4212
      : : : : : : : : : : : : : : :
Qy 351 GlyThrGlyGlnAlaThrSerArgAsn-----GlyHisSerAlaArgGlnHisValVal 368
      ||| : : : : : ||| : : : : :
Db 4213 GCTACCGGTGAGACAAACAACCAATACTTTAGCTGTGAACGACTACCATACTGTAGCT 4272
      : : : : : : : : : : : : : : :
Qy 369 AlaAspThr 371
      : : : : : ||| : : : : :
Db 4273 GCTGAGACG 4281
```

Search completed: October 30, 2002, 10:50:26

Job time : 68.8495 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:04:33 ; Search time 169.922 Seconds
(without alignments)
3768.837 Million cell updates/sec

Title: US-09-836-602-2

Perfect score: 1895

Sequence: 1 MPRGWTALCLLSLLPSGFM.....QATSRNGHSARQHVADTEL 373

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xl
-O=cgcn2_1/USPTO.spool/US09836602/runat_29102002_101143_3758/app.query.fasta_1.1429
-DB=N.Geneseq_032802 -QFMT=tastep -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=Dlosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602.ecgn.1.1.312.eRunat_29102002_101143_3758 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq_032802:.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1895	100.0	1122	24	AAD22661	Human full-length Nucleotide sequenc
2	1895	100.0	2463	22	AAH75132	Human CD-34 polynu
3	1891	99.8	2615	21	AAF20898	Human adenosine re
4	1891	99.8	2615	21	AAA34776	Human CD-34 polynu
5	1891	99.8	3490	21	AAF20899	Human adenosine re
6	1891	99.8	3490	21	AAA34777	Human CD34 surface
7	1574	83.1	951	24	AAD22662	Human CD34 surface
8	1509	79.6	906	24	AAD22663	Porcine CD34 from
9	904	47.7	1065	18	AAT59508	Human secreted pro
10	508	26.8	615	21	AAC00249	Human breast cell
11	311	16.4	233	22	ABA51134	Human foetal liver
12	311	16.4	233	22	ABA69132	Probe #14530 for g
13	311	16.4	233	22	ABA36064	Human brain expres
14	311	16.4	233	22	AAK17442	Human bone marrow
15	311	16.4	233	22	AAK43242	Probe #13951 for g
16	311	16.4	233	22	AAI24018	Probe #18003 used t
17	311	16.4	233	22	AAI49317	Human breast cell
18	311	16.4	233	22	AAI09602	Human breast cell
19	265	14.0	209	22	ABA50740	Human foetal liver
20	265	14.0	209	22	ABA68710	Probe #14140 for g
21	265	14.0	209	22	ABA35674	Human brain expres
22	265	14.0	209	22	AAK17050	Human bone marrow
23	265	14.0	209	22	AAK42833	Probe #13530 for g
24	265	14.0	209	22	AAI23597	Probe #17596 used
25	265	14.0	209	22	AAI48910	Probe #9203 used t
26	265	14.0	209	22	AAI09212	Novel human diagno
27	233	12.3	235	22	AAS39236	Human breast cell
28	174	9.2	406	22	ABA45618	Human foetal liver
29	174	9.2	406	22	ABA56124	Probe #4241 for ge
30	174	9.2	406	22	AAK25775	Human brain expres
31	174	9.2	406	22	AAK04311	Human bone marrow
32	174	9.2	406	22	AAK29806	Probe #4329 for ge
33	174	9.2	406	22	AAI14396	Probe #4456 used t
34	174	9.2	406	22	AAI35770	Probe #4210 used t
35	174	9.2	406	22	AAI04219	Enterococcus faeca
36	159	8.4	13154	20	AAI13275	Murine PCPL codin
37	152	8.0	2131	22	AAH25852	Drosophila melanog
38	150.5	7.9	2336	23	ABL25662	Human DNA sequence
39	148.5	7.8	5848	24	AAS94835	Insert from cosmid
40	148	7.8	40875	18	AAT80043	Genomic DNA encodi
41	147.5	7.8	7720	21	AAAS3800	Drosophila melanog
42	146	7.7	4116	23	ABL24389	Human kidney cell
43	146	7.7	5894	22	AAH57511	Drosophila melanog
44	146	7.7	6116	23	ABL24388	Staphylococcus aur
45	146	7.7	7035	23	AAS54978	

ALIGNMENTS

RESULT 1

AAD22661

ID AAD22661 standard; DNA; 1122 BP.

XX AAD22661;

AC AAD22661 (first entry)

DT 26-FEB-2002

XX Human full-length CD34 (fLCD34) surface antigen encoding DNA.

DE Human; surface marker; surface antigen; T lymphocyte; gene therapy;

XX full-length CD34; fLCD34; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..1122

FT /*tag= a

FT XX /product= "Human full-length CD34 surface antigen"
 PN EPI148066-A1.
 XX 24-OCT-2001.
 PD 18-APR-2001; 2001EP-0109374.
 PF 18-APR-2000; 2000DE-1019075.
 PR (ZAND/) ZANDER A R.
 PA zander AR;
 XX WPI: 2002-019289/03.
 PI P-PSDB; AAE13541.
 DR
 XX New gene transfer vector (accession number DSM13396) containing a
 PT transgene and a nucleic acid sequence coding for a surface marker,
 PT useful in gene therapy, and for detecting genetically modified cells or
 PT cells which do not express CD34 -
 XX
 PS Claim 3; Page 9-11; 28pp; English.
 XX The patent discloses a gene transfer vector (accession number DSM13396)
 CC containing a transgene and a nucleic acid sequence coding for a surface
 CC marker. The surface marker is the CD34 surface antigen, its fragment or
 CC variant. The vector is useful for in vitro transduction of T lymphocytes,
 CC for gene therapy, and in the enrichment, detection and analysis of cells
 CC in vitro that do not naturally express CD34. T lymphocytes transduced
 CC with the vector are also useful in gene therapy. The CD34 nucleic acid
 CC sequences (marker genes), their fragments or variants are used for
 CC detecting genetically modified cells or cells which do not naturally
 CC express CD34. The present sequence is a DNA encoding human full-length
 CC CD34 (fLCD34) surface antigen.
 XX
 SQ Sequence 1122 BP; 305 A; 311 C; 270 G; 236 T; 0 other;

 Alignment Scores:
 Pred. No.: 6,89e-138 Length: 1122
 Score: 1895.00 Matches: 373
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0

 US-09-836-602-2 (1-373) x AND22661 (1-1122)

 Qy 1 MetProArgGlyThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 DB 1 ATGCCGGGGCTGGACCGCGCTTGTGCTGAGTTTGTCTTCTGGCTTCATCAGT 60

 Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 DB 61 CTTGACAAACAGGTAAGTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 120

 Qy 41 SerThrAsnValSerTyrglnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 DB 121 TCTACAANTGTATCTACCAAGAACTACAAACCTAGTACCTTGGAGTACCAGCGCTG 180

 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVallys 80
 DB 181 CACCCTGTGTCACATGCAATGCAATGAGGCCACAAACATCACAGAAACGACAGTCAA 240

 Qy 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
 DB 241 TTCACATCTACCTCTGTGAACCTCAGTTATGGAAACACAAACTCTCTGTGCCAGTCA 300

 Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 DB 301 CAGACCTCTGTATCAGCAGAGTGTCCACCACCCCAACGTTTCAACTCCAGAGACA 360

 Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrSerThrSer 140

DB 361 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCACAGC 420
 QY 141 LeuAlaThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIleLys 160
 DB 421 CTTGCAACATCTCCCACTAAACCTATACATCATCTCTCTCTATCTCCCTAAAGTGACATCAAG 480
 QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 DB 481 GCAGAAATCAANTGTTTCAGGCATCAGAGAAAGTGAATGACTCAGGGCATCTGCCTGGAG 540
 QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluLeuAlaArg 200
 DB 541 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAGAAGACACAGGGAGAGGCGCTGGCCCGA 600
 QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 DB 601 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGCTGGGGCCCAAGTATGCTCCCTGCTC 660
 QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
 DB 661 CTTGCCCACTGTGAGGTGAGGCCCTCAGTGCTACTGCTGCTGTGGCCCAACAGACAGAA 720
 QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
 DB 721 ATTTCCAGCAAACTCCAACTTATGAAAAGACCAATCTGACCTGAAAAAGTGGGATC 780
 QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrrSerGlnLysThrLeuIle 280
 DB 781 CTAGATTTCATGACGCAAGATGTGCAAGCCACAGAGCTATTCCCAAAAGACCCCTGATT 840
 QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrrPheLeuMet 300
 DB 841 GCACCTGCTACCTCGGAGGCCCTGCTGCTGCTGCTGGGCATCACTGGCTATTTCCTGATG 900
 QY 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrrTyrrThrGlu 320
 DB 901 AATCGCGCAGCTGGAGCCCAACAGAGAGAAAGGCTGGCGAAGACCCCTATTACACGGAA 960
 QY 321 AsnGlyGlyGlnGlyTyrrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
 DB 961 AACGGTGGAGCCAGGCTATAGCTCAGGACCTGGGACCTCCCTCGAGGCTCAGGGAAAG 1020
 QY 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
 DB 1021 GCCAGTGTGAACCGAGGGCTCAGGAAACGGGACCGCCAGGCCACCTCCAGAAACGGC 1080

 QY 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
 DB 1081 CATTCAGCAAGACAAACACGTTGGTGGCTGATACCGAATTG 1119

 RESULT 2
 AAH75132 standard; DNA; 2463 BP.
 ID AAH75132
 XX AC
 XX AAH75132;
 XX 13-NOV-2001 (first entry)
 XX Nucleotide sequence of a human CD34 polypeptide.
 DE Human; CD34 gene; blast.crisis; chronic myelogenous leukemia;
 KW nm23-H4 kinase gene; ss.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FT CDS 91..1248
 FT /*tag= a
 FT /product= "CD34"
 XX WO200164946-A1.
 PN XX

PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-JP01485.
XX
PR 02-MAR-2000; 2000JP-0058043.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Mano H, Miyazato A, Ueno S, Yoshida K, Yamanaka T, Ikeda U;
PI Shimada K, Hatake K, Ozawa K, Asada K, Kato I;
XX
XX WPI; 2001-550191/61.
DR P-PSDB; AAG67120.
XX
PT Method for detecting chronic myelogenous leukemia by comparing
PT expression levels of CD34 and nm23-H4 genes -
XX
PS Disclosure; Page 36-40; 60pp; Japanese.
XX
CC The present sequence encodes a human CD34 polypeptide. The
CC specification describes a method of detecting blast crisis in chronic
CC myelogenous leukemia. The method comprises comparing the amounts of
CC expression of at least two genes in a sample, particularly CD34 gene
CC and nm23-H4 kinase gene. The method allows the worsening stages of
CC chronic myelogenous leukemia to be easily detected at a high
CC reliability.
XX
SQ Sequence 2463 BP; 589 A; 709 C; 580 G; 585 T; 0 other;

Alignment Scores:
Pred. No.: 1.96e-137 Length: 2463
Score: 1895.00 Matches: 373
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-2 (1-373) x AAH75132 (1-2463)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
DB 127 ATGCCGGGGCTGGACCGCGCTTGGCTGCTGAGTTGCTGCCCTCTGGGGTTCATGAGT 186

QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 187 CTTGACACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 246

QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 247 TCTACAAATGTATCTACCAAGAACTACACACCTAGTACCTTGGAAAGTACCAGCCTG 306

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
DB 307 CACCCTGTGTCACATGGCAATGAGCCACCAACAAACATCACAGAAACGACAGTCAA 366

QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAspSerSerValGlnSer 100
DB 367 TTCACATCTACCTCTGTGATACCTCAGTTTATGGAACACAAACTCTCTCTCCAGTCA 426

QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 427 CAGACCTCTGTAATCAGCACAGTGTCCACCACCCCAACAGTTCACCTCCAGAGACA 486

QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
DB 487 ACCTTTGAAGCCTTAGCTGTACCTGGAATGTTTTCAGACCTTTCAACCACTAGCATAGC 546

QY 141 LeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIleLys 160
DB 547 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTTCTCTATCTTAAGTGACATCAAG 606

QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180

DB 607 GCAGAAATCAATGTTTCAGGCATCAGAGAAGTGAATTTACTCAGGGCATCTGCCTGGAG 666

QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200

DB 667 CAAAATAAGACCTCCAGCTGTCGGAGTTTAAAGAGGACAGGGAGAGGGCCTGCCCCGA 726

QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220

DB 727 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCCTC 786

QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240

DB 787 CTTGCCAGTCTGAGGTGAGGCTCAGTGTCTACTGCTGTGGCCAAACAGACAGAA 846

QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLysLysLeuGlyIle 260

DB 847 ATTTCCAGCAACTCCAATTTATGAAGAAGCACCAATCTGACCTGAAAAGCTGGGGATC 906

QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280

DB 907 CTGATTTTCACTGAGCAAGATGTTGCAAGCCACAGAGCTATTCCAAAAGACCCCTGATT 966

QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300

DB 967 GCACGTGGTCACTCGGAGCCCTGCTGCTGTGGGCATCCTGCGCTATTTCCTGATG 1026

QY 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrThrGlu 320

DB 1027 AATCCCGCAGCTGGAGCCCAACAGAGAGAAAGCTGGGCAAGACCTTTATTACAGGAA 1086

QY 321 AsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340

DB 1087 AACGCTGGAGGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTGAGGCTCAGGGAAG 1146

QY 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGluAlaThrSerArgAsnGly 360

DB 1147 GCCAGTGTGAACCGAGGGGCTCAGGAAAACGGGACCGGCCACCTCCAGAAACGCG 1206

QY 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373

DB 1207 CATTCAGCAAGACAACACGTGCTGCTGATACCGAATTG 1245

RESULT 3
AAF20898
ID AAF20898 standard; DNA; 2615 BP.
XX
AC AAF20898;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human CD-34 polynucleotide fragment #2465.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antialsthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX

PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.

XX NYCE JW;

XX WPI; 2000-679539/66.

XX Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -

XX Disclosure; Page 292-293; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytoskeletal activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (ARDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

XX SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:

Pred. No.: 4.32e-137 Length: 2615
 Score: 1891.00 Matches: 372
 Percent Similarity: 100.00% Conservatives: 1
 Best Local Similarity: 99.73% Mismatches: 0
 Query Match: 99.79% Indels: 0
 DB: 21 Gaps: 0

US-09-836-602-2 (1-373) x AAF20898 (1-2615)

Qy 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 294 ATGCCCGGGGCTGGACCGGCTTTGCTGAGTTTGTCTTCTGGGTTTCATGAT 353
 Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProGluThrGlnGlyThrPheSerAsnVal 40
 Db 354 CTTGACAAACACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 413
 Qy 41 SerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 414 TCTACAATGTATCTTACCAAGAAACTACACACCTAGTACCTTGGAGTACCGGCTG 473
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
 Db 474 CACCCCTGTGCTCAACATGGCAATGAGGCCACACAAACATCACAGAAGACAGTCAAA 533
 Qy 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100

Db 534 TTCACATCTACCTCTGTGATAAACCTCAGTTTATGGAAACACAAACTCTTCTGTCCAGTCA 593
 Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 Db 594 CAGACCTCTGTATATCAGCAGAGTGTTCACCACCCACCCACACGTTTCAACTCCAGAGACA 653
 Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 Db 654 ACCTTGAAGCTACGCTGTACCTGGAAATGTTTCAGACCTTTCACCACTAGCAGTAC 713
 Qy 141 LeuAlaThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIleLys 160
 Db 714 CTTGCAACATCTCCACTAAACCTTATACATCATCTCTCTCTATCTTAAAGTGACATCAAG 773
 Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 Db 774 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCTGGAG 833
 Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 Db 834 CAAATAAGACCTCCACCTGTGGGAGCTTTAAGAAGACAGAGGGGAGGCGCTGGCCCGA 893
 Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 Db 894 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGCTGTGGGCCAGGATGCTCCCTGCTC 953
 Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
 Db 954 CTTGCCAGCTGTAGGTGAGGCTCAGTGCTACTGTGTGTGGTCTTGGGCCACAGAACAGAA 1013
 Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
 Db 1014 ATTTCCAGCAACTCCAACTTATGAAAGACCACTGACCTGAAAGAGCTGGGATC 1073
 Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyThrSerGlnLysThrLeuIle 280
 Db 1074 CTAGATTTCACTAGCAAGATGTTGCAAGCCACCCAGAGCTATTCCCAAAAGACCTGATT 1133
 Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyThrPheLeuMet 300
 Db 1134 GCATGTGCTACCTCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1193
 Qy 301 AsnArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyThrTyThrGlu 320
 Db 1194 AATCGCGCAGCTGGAGCCGCCACAGAGGAAAGGCTGGCGAAGACCTTATTACACGGAA 1253
 Qy 321 AsnGlyGlyGlnGlyTyThrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
 Db 1254 AACGGTGGAGGCCAGGCTATAGCTCAGGACCTGGGACCTCCCTGAGGCTCAGGGGAAAG 1313
 Qy 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
 Db 1314 GCCAGTGTGAACCGAGGGGCTCAGAAAACGGGACCGCCAGCCACCTCCAGAAACGGC 1373
 Qy 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
 Db 1374 CATTCAGCAAGACAAACACGTGGTGGCTGATACCGAATTG 1412
 RESULT 4
 AAA34776
 ID AAA34776 standard; DNA; 2615 BP.
 XX
 AC AAA34776;
 XX
 DT 28-JUL-2000 (first entry)
 XX
 DE Human adenosine receptor related polynucleotide SEQ ID NO:2465.
 XX
 KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytoskeletal; pulmonary vasoconstriction; asthma;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;

KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX Homo sapiens.
XX WO200009525-A2.
XX PD 24-FEB-2000.
XX PF 03-AUG-1999; 99WO-US17712.
XX PR 03-AUG-1998; 98US-0095212.
XX PA (UYEC-) UNIV EAST CAROLINA.
XX PI Nyce JW;
XX DR WPI; 2000-205971/18.
XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers
PS Disclosure: Page 602; 1343pp; English.
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA3992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:
Pred. No.: 4.32e-137 Length: 2615
Score: 1891.00 Matches: 372
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.73% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-2 (1-373) x AAA34776 (1-2615)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
DB 294 ATGCGCGGGGTGGACCGCGTGTGCTTGTGAGTTTCTGCCCTCTCGGTTCATGAGT 353

QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 354 CTTGACACACGGTACTGCTACCCAGAGATTACCTACCAGGGGAACATTTTCAATGTT 413

QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 414 TCTACAAATGTATCTCTACCAAGAAACTACACACACCTAGTAGCTTGAAGTAGTACCGCTG 473

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
DB 474 CACCCCTGTCTCAACATGGCAATGAGGCCACAAACAAACATCACAGAAACGACAGTCAA 533

QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 534 TTCACATCTACCTCTGTGATAACCTCAGTTATGGAAACACAAACTCTTCTGTCCAGTCA 593

QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 594 CAGACCTCTGTAATCAGCACAGTGTTCACACCCAGCCCAACGTTTCAACTCCAGAGACA 653

QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
DB 654 ACCTTTGAAGCCTAGCCTGTCACTGGAAATGTTTCAGACCTTTCAACCACTAGCATAAGC 713

QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
DB 714 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCCTATCCTAAGTAGCATCAAG 773

QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
DB 774 GCAGAAATCAATGTTTCAGGCATCAGAGAAGTGAATTCACTCAGGGCATCTGCCTGGAG 833

QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
DB 834 CAAATAAAGACCTCCACAGCTGTGCGAGTTTAAAGAACGACAGGGAGAGGGCCTGGCCCA 893

QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyValAlaGlnValCysSerLeuLeu 220
DB 894 GTGCTGTGTGGGAGGAGGAGCGCTGATGCTGATGTGGGGCCCGCAGGTATGCTCCCTGCTC 953

QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
DB 954 CTTGCCAGCTCTGAGGTGAGGCTCAGTCTACTGCTGCTGTGGTCTTGGCCCAACAGACAGAA 1013

QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
DB 1014 ATTTCCAGCAAACTCCAACTTATGAAAAAGCACCAATCTGACCTGAAAAAGCTGGGGATC 1073

QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
DB 1074 CTAGATTTCTACTGAGCAAGATGTTGCAAGCCACAGAGCTATTTCCAAAAGACCTTGATT 1133

QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
DB 1134 GCACGTGTCACCTCGGGAGCCCTGCTGGCTGCTTGGGCATCACTGGCTATTTTCTCTGATG 1193

QY 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrThrThrGlu 320
DB 1194 AATCGCGCGAGCTGGAGCCCCACAGAGAAAGGCTGGGGAGAGACCTTATTATACAGGAA 1253

QY 321 AsnGlyGlyGlnGlnTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
DB 1254 AACGGTGGAGGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTCCTCAGGGCTCAGGAAAG 1313

QY 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
DB 1314 GCCAGTGTGAACCGAGGGGCTCAGAAAAACGGGACCGGCGGACGCCACCTCCAGAAACGGC 1373

QY 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
DB 1374 CATTCAGCAAGACACACGCTGGTGTGCTGATACGAATTG 1412

RESULT 5
AAF20899
ID AAF20899 standard; DNA; 3490 BP.
XX
AC AAF20899;

XX 14-MAR-2001 (first entry)
XX Human CD-34 polynucleotide fragment #2466.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antilasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
XX WO200062736-A2.
XX
XX 26-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-US08020.
XX
XX 06-APR-1999; 99US-0127958.
XX (UVEC-) UNIV EAST CAROLINA.
XX (NYCE/) NYCE J W.
XX
XX Nyce JW;
XX
XX WPI; 2000-679535/66.
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
XX trigger adenosine receptors during metabolism, useful e.g. for treating
XX cancers and respiratory obstructions -
XX
XX Disclosure; Page 291-292; 1592pp; English.
XX
XX The present invention describes low adenosine (A) content antisense
XX oligonucleotides and compositions (I) comprising them. In the antisense
XX oligonucleotides the A is replaced by a 'Universal' or alternative base.
XX (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
XX immunosuppressive, antilasthmatic, hypotensive and cytostatic activities.
XX The antisense oligonucleotides and (I) can be used to down-regulate the
XX expression and or activity of target polypeptides associated with the
XX lung/respiratory disorders and malignancies, such as stimulating and
XX activating peptide factors and transmitters, transcription factors,
XX immunoglobulins and antibodies, antibody receptors, cytokines and
XX chemokines, endogenously produced specific and non-specific enzymes,
XX binding proteins, adhesion molecules and their receptors, cytokine and
XX chemokine receptors, adenosine receptors, bradykinin receptors, central
XX nervous system (CNS) and peripheral nervous and non-nervous system
XX receptors, CNS and peripheral nervous and non-nervous system peptide
XX transmitters, defensins, growth factors, vasoactive peptides and
XX receptors, binding proteins and malignancy associated proteins. The
XX antisense oligonucleotides may be used in this way to treat disorders
XX including respiratory obstruction (especially pulmonary obstruction
XX and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
XX and/or surfactant hypoproduction which are associated with a disease or
XX condition selected from pulmonary vasoconstriction, inflammation,
XX allergies, asthma, impeded respiration, respiratory distress syndrome
XX (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
XX hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
XX pulmonary transplantation rejection, pulmonary infections, bronchitis,
XX and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
XX fragments and antisense oligonucleotides used in the exemplification of
XX the present invention.
XX
XX Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 6.34e-137 Length: 3490

Score: 1891.00 Matches: 372
Percent Similarity: 100.00% Conservativity: 1
Best Local Similarity: 99.73% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 21 Gaps: 0
US-09-836-602-2 (1-373) x AAF20899 (1-3490)
Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1169 ATGCCCGGGGCTGGACCGCGCTTGTGCTGAGTTTGCTTGGCTTCATGAGT 1228
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 1229 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAATGTT 1288
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 1289 TCTACAAATGTATCTTACCAAGAACTACAACACCTAGTACCTTGGAAAGTACCAGCCTG 1348
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 1349 CACCTGTGCTCAACATGGCAANTGAGGCCAACAAACATCACAGAAACACAGTCAAA 1408
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 1409 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTGTCTCAGTCA 1468
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 1469 CAGACCTCTGTAATCAGCACAGTGTTCACACCCCGCCACGTTTCAACTCCAGAGACA 1528
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 1529 ACCTTGAAGCCCTAGCCTGTCCACCTGGAAATGTTTACAGACCTTCAACCACTAGCACTAGC 1588
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 1589 CTTGCAACATCTCCACTAAACCCCTATATACATCATCTTCTCTATCTAAGTGACATCAAG 1648
Qy 161 AlaGluIleLysCysSerGlyIleArgGluVallysLeuThrGlnGlyIleCysLeuGlu 180
Db 1649 GCAGAAATCAATGTTCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCTGCTGGAG 1708
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 1709 CAATTAAGACCTCCAGCTGTCCGAGTTTAAAGAGACAGAGGGAGGGCTGGCCCGA 1768
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 1769 GTGCTGTGGGAGGAGCAGGCTGATGCTGCTGGGGCCAGGTATGCTCCTCGTCTC 1828
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
Db 1829 CTTGCCAGTCTGAGGTGAGGCCCTCAGTGTCTACTGCTGTGCTTGGCCCAACAGACAGAA 1888
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLysGlyIle 260
Db 1889 ATTTCCAGCAACCTCCAACTTATGAAAAGCACCCTATGACCTGAAAAGACTGGGGATC 1948
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1949 CTAGATTTCAGTACGAGAGATGTTGCAAGCCACAGAGCTATTTCCCAAGACACCTGATT 2008
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 2009 GCATGTGTCACCTCGGGAGCCCTGCTGGCTGCTTGGGCATCACTGGCTATTTCCTGATG 2068
Qy 301 AsnArgArgSerTyrProThrGlyClnuArgLeuGlyGluAspProTyrTyrThrGlu 320
Db 2069 AATGCCGACAGCTGGAGCCCCACAGGAAAGGCTGGCGAAGACCCCTTATTACAGGAA 2128
Qy 321 AsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340

|||||
Db 2129 AACGGTGAGCGCAGGGCTATAGCTCAGGACCTCGGACCTCCCTCGAGGCTCAGGGAAG 2188
Qy 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
Db 2189 GCCAGTGTGAACCGAGGGGCTCAGAAAAACGGGACCGCCAGGCCACCTCCAGAAACGGC 2248
Qy 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 2249 CATTGAGCAAGACACACGCTGGTGGCTGATACCGAATTG 2287
RESULT 6
AAA34777
ID AAA34777 standard; DNA; 3490 BP.
XX
XX
AC AAA34777;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2466.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
XX WO200009525-A2.
XX
XX PD 24-FEB-2000.
XX
XX PF 03-AUG-1999; 99WO-US17712.
XX
XX PR 03-AUG-1998; 98US-0095212.
XX
XX PA (UYEC-) UNIV EAST CAROLINA.
XX
XX PI Nyce JW;
XX
XX DR WPI; 2000-205971/18.
XX
XX PT New antisense oligonucleotides useful for treating e.g. pulmonary
XX vasoconstriction, inflammation, allergies, asthma, hypertension,
XX bronchitis, emphysema, respiratory distress syndrome, ischemia or
XX cancers -
XX
XX PS Disclosure; Page 603; 1343pp; English.
XX
XX CC The present invention describes a new composition comprising an
XX antisense oligonucleotide (ON) with low adenosine (up to 15%), which
XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
XX inflammation. The ON can have antiinflammatory, antiallergic,
XX antiasthmatic, cytostatic and analgesic activities. The compositions are
XX useful for the treatment of diseases associated with inflammation,
XX impaired airways, including lung disease and diseases whose secondary
XX effects afflict the lungs of a subject. They can be used for treating
XX e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
XX asthma, impeded respiration, respiratory distress syndrome, pain, cystic
XX fibrosis, pulmonary hypertension, emphysema, chronic obstructive
XX pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
XX carcinomas, and cancers which may metastasize to the lungs, including
XX breast and prostate cancer. The reduction of the adenosine content of
XX the ONs reduces side effects. The A-containing ONs break down with the
XX release of deoxyadenosine which activates adenosine receptors causing
XX bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
XX nucleotide sequences given in the sequence listing from the present
XX invention, which correspond to SEQ ID NO:1 to 2815, and then the last
XX 185 sequences are also called SEQ ID NO:1 to 185, but the sequences

CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
XX given in the sequence listing.
SQ Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;
Alignment Scores:
Pred. NO.: 6.34e-137 Length: 3490
Score: 1891.00 Matches: 372
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.73% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 21 Gaps: 0
US-09-836-602-2 (1-373) x AAA34777 (1-3490)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1169 ATGCCGCGGGCTGGACCGCGCTTGTCTGCTGAGTTGCTGCCCTCTGGGTTCAGT 1228
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 1229 CTTGACAAACAGCGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 1288
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 1289 TCTCAAAATGATCTCTACCAAGAACTACACACCTAGTACCTTGGAGTAGTACCGCTCG 1348
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 1349 CACCCTGTGTCTCAACATGGCAATGAGGCCACACAAACATCACAGAACGACAGTCAA 1408
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 1409 TTCACATCTACCTCTGTGTATACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 1468
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 1469 CAGACCTCTGTAAATCAGCACAGTGTTCACCCAGCCAGCCAACTTCACTCCAGAGACA 1528
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 1529 ACCTTGAAGCCTAGCCTGTCACTCGAAATGTTTTCAGACCTTTCAACCACTAGCACTAGC 1588
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 1589 CTTGCAACATCTCCCACTAAACCTATATCATCATCTCTCTCTATCTTAAGTACATCAAG 1648
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 1649 GCAGAAATCAATGTTTTCAGGCATCAGAGAAAGTGAATTCAGCTCAGGGCATCTGCTGGAG 1708
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyGlyLeuAlaArg 200
Db 1709 CAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAGGACAGGGAGAGGCGCTGGCCCA 1768
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 1769 GTGCTGTGTGGGAGGAGCAGCTGATGCTGATGTGGGGCCAGGTATGCTCCCTGCTC 1828
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 1829 CTTGCCAGTCTGAGGTGAGGCTCAGTGTCTACTGTCTTGGCCCAACAGACAGAA 1888
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGlyIle 260
Db 1889 ATTTCCAGCAAACTCCAACTTATGAAAGACCAATCTGACCTGAAAGAGCTGGGGATC 1948
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuLe 280
Db 1949 CTAGATTTTCACTGAGCAAGATGTTCAGAGCCAGAGCTATTCCCAAGAACACCTGATT 2008

QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
 |||||
 Db 2009 GCACCTGGTCACCTCGGAGCCCTGGCTGTCTTGGGCATCACTGGCTATTTCCTGATG 2068
 QY 301 AsnArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrThrGlu 320
 |||||
 Db 2069 AATCGCGCAGCTGGAGCCGCCACAGAGAAAGGCTGGGGAGACCTTATTACAGGAA 2128
 QY 321 AsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
 |||||
 Db 2129 AACGGTGGAGGCCAGGCGCTATAGCTCAGGACCTGGGACCTGCCCTGAGGCTCAGGGAAG 2188
 QY 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
 |||||
 Db 2189 GCCAGTGTGAACCGAGGGCTCAGAAAAGGGAGCGCCAGCGCCACCTCCACAAACGGC 2248
 QY 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
 |||||
 Db 2249 CATTCAGCAAGACACACGCTGGTGGCTGATACCGAATTG 2287

RESULT 7

AAD22662

ID AAD22662 standard; DNA; 951 BP.

XX AC AAD22662;

DT 26-FEB-2002 (first entry)

XX DE Human CD34 surface antigen truncated variant (tCD34) encoding DNA.

XX KW Human; surface marker; surface antigen; T lymphocyte; gene therapy;

XX KW CD34 truncated variant; tCD34; ds.

XX OS Homo sapiens.

XX OS Synthetic.

XX Key Location/Qualifiers

FH 1..951

FT /tag- a

FT /product- "Human CD34 surface antigen

FT truncated variant (tCD34)"

XX EP1148066-A1.

XX PD 24-OCT-2001.

XX PF 18-APR-2001; 2001EP-0109374.

XX PR 18-APR-2000; 2000DE-1019075.

XX PA (ZAND/) ZANDER A R.

XX Zander AR;

XX WPI; 2002-019289/03.

XX P-PSDB; AAE13542.

XX PT New gene transfer vector (accession number DSM13396) containing a

XX PT transgene and a nucleic acid sequence coding for a surface marker,

XX PT useful in gene therapy, and for detecting genetically modified cells or

XX PT cells which do not express CD34.

XX PS Claim 3; Page 12-13; 28pp; English.

XX CC The patent discloses a gene transfer vector (accession number DSM13396)

XX CC containing a transgene and a nucleic acid sequence coding for a surface

XX CC marker. The surface marker is the CD34 surface antigen, its fragment or

XX CC variant. The vector is useful for in vitro transduction of T lymphocytes,

XX CC for gene therapy, and in the enrichment, detection and analysis of cells

XX CC in vitro that do not naturally express CD34. T lymphocytes transduced

XX CC with the vector are also useful in gene therapy. The CD34 nucleic acid

XX CC sequences (marker genes), their fragments or variants are used for

CC detecting genetically modified cells or cells which do not naturally
 CC express CD34. The present sequence is a DNA encoding human CD34 surface
 CC antigen truncated variant (tCD34).

XX Sequence 951 BP; 261 A; 264 C; 213 G; 213 T; 0 other;

Alignment Scores:

Pred. No.: 4,26e-113 Length: 951
 Score: 1574.00 Matches: 313
 Percent Similarity: 99.37% Conservative: 1
 Best Local Similarity: 99.05% Mismatches: 2
 Query Match: 83.06% Indels: 0
 DB: 24 Gaps: 0

US-09-836-602-2 (1-373) x AAD22662 (1-951)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 |||||
 Db 1 ATGCCCGGGGCTGGAGCCGCTTGTCTGTGAGTTGCTGCTTCTGGGTTTCATGAGT 60
 QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 |||||
 Db 61 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGAGACATTTTCAATGTT 120
 QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 |||||
 Db 121 TCTACAAATGTATCTTACCAAGAAACTACAAACACCTAGTACCTTGGAGTACGAGCTG 180
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
 |||||
 Db 181 CACCCCTGTGTTCAACATGGCAATGAGGCCACCAACAACATCACAGAAAGCAGAGTCAA 240
 QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 |||||
 Db 241 TTCACATCTACCTCTCTGTGATAACCTCAGTTTATGGAACAACAACACTCTCTGTCCAGTCA 300
 QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 |||||
 Db 301 CAGACCTCTGTAATCAGCACAGTGTTCACACCCCAACGTTTCACTCCAGAGACA 360
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 |||||
 Db 361 ACCTTGAAGCCTAGCCTGTCCCTGGAAATGTTTCAGACCTTTCAACACCTAGCCTAGC 420
 QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
 |||||
 Db 421 CTTGCAACATCTCCACTAACCTTATACATCATCTTCTCTATCTTAACTAGTACATCAAG 480
 QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 |||||
 Db 481 GCAGAAATCAATGTTTCAGGCATCAGAGAAAGTGAATGTGACTCAGGGCATCTGCTGGAG 540
 QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 |||||
 Db 541 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGGGAGAGGGCCTGGCCCGA 600
 QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 |||||
 Db 601 GTCTGTGTGGGAGGAGCAGGCTGATGCTGGGCCCCAGGATGCTCCCTGCTC 660
 QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
 |||||
 Db 661 CTTGCCCCAGTCTGAGGTGAGGCTCTAGTGTCTACTGTGCTTGTGGCAACAGAACAGAA 720
 QY 241 IleSerSerLysLeuGlnLeuMetLysHisGlnSerAspLeuLysLysLeuGlyLe 260
 |||||
 Db 721 ATTTCAGCAAACTCCAACTTATGAAAAGACCAACTCTGACCTGAAAAGCTGGGGATC 780
 QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
 |||||
 Db 781 CTAGATTTCAGTACGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAAGACCTGATT 840
 QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300

Db 841 GCACCTGGTCACTCGGGAGCCCTGCTGGCTGTCTTGGGCATCAGTGGCTATTTCCTGATG 900
QY 301 AsnArgSerTrpSerProThrGlyGluArgLeuGlyGluAspPro 316
Db 901 AATCGCCGAGCTGGAGCCCCACAGGAGAAAGGCTGGAAGCTAGAACCA 948
RESULT 8
ID AAD22663
XX AAD22663 standard; DNA; 906 BP.
AC AAD22663;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human CD34 surface antigen deleted variant (dCD34) encoding DNA.
KW Human; surface marker; surface antigen; T lymphocyte; gene therapy;
KW CD34 deleted variant; dCD34; ds.
XX Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
CDS 1..906
FT /*tag= a
FT /product= "Human CD34 surface antigen
FT deleted variant (dCD34)"
XX
PN EP1148066-A1.
XX
PD 24-OCT-2001.
XX
PF 18-APR-2001; 2001EP-0109374.
XX
PR 18-APR-2000; 2000DE-1019075.
XX
PA (ZAND/) ZANDER A R.
XX
PI Zander AR;
XX
DR WPI; 2002-019289/03.
DR P-PSDB; AAEE13543.
XX
XX New gene transfer vector (accession number DMS13396) containing a
PT transgene and a nucleic acid sequence coding for a surface marker,
PT useful in gene therapy, and for detecting genetically modified cells or
PT cells which do not express CD34 -
XX
PS Claim 3; Page 14-15; 28pp; English.
XX
CC The patent discloses a gene transfer vector (accession number DMS13396)
CC containing a transgene and a nucleic acid sequence coding for a surface
CC marker. The surface marker is the CD34 surface antigen, its fragment or
CC variant. The vector is useful for in vitro transduction of T lymphocytes,
CC for gene therapy, and in the enrichment, detection and analysis of cells
CC in vitro that do not naturally express CD34. T lymphocytes transduced
CC with the vector are also useful in gene therapy. The CD34 nucleic acid
CC sequences (marker genes), their fragments or variants are used for
CC detecting genetically modified cells or cells which do not naturally
CC express CD34. The present sequence is a DNA encoding human CD34 surface
CC antigen deleted variant (dCD34).
XX
SQ Sequence 906 BP; 247 A; 250 C; 199 G; 210 T; 0 other;

Alignment Scores:
Pred. No.: 4.37e-108 Length: 906
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 79.63% Indels: 0
DB: 24 Gaps: 0

US-09-836-602-2 (1-373) x AAD22663 (1-906)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1 ATGCCCGGGGCTGGACCGGCTTCTGCTGAGTTTGTCTGCTTCTGGTTTCATGAGT 60
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 61 CTTGACAACAAGGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 120
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 121 TCTACAAATGTATCTTACCAAGAACTACAACACCTAGTACCTTGGAAAGTACCAAGCTG 180
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 181 CACCCTGTCTCTACATGGCAATGAGGCCACAAACAACATCACAGAAACGACAGTCAA 240
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 241 TTCACATCTACCTCTCTGTATACCTCAGTTTATGGAACACACAAACTCTTCTGCTCCAGTCA 300
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 301 CAGACCTCTGTATATCAGCACAGTGTTCACACCCCAACGCTTTCACATCCAGAGACA 360
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 361 ACCTTGAAGCCTAGCCTGTCACTGGAATGTTTCAGACCTTTCACACCTACACTAGCTAGC 420
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 421 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTCTAAGTGACATCAAG 480
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 481 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGGCATCTGCCTGGAG 540
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 541 CAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAGACAGGGGAGAGGGCTTGGCCCGA 600
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlnValCysSerLeuLeu 220
Db 601 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGTGCTGGGCCCCAGGTATGCTCCCTGCTC 660
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 661 CTTGCCACGTCTCAGGTGAGGCTCAGTGTCTACTGTGCTTGGCCAAACAGAACAGAA 720
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGlyIle 260
Db 721 ATTTCAGCAAACTCCAACTTATGAAAGACCACTGACCTGAAAGCTGGGGATC 780
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 781 CTAGATTCTCACTCAGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAGACCTGATT 840
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 841 GCACCTGTCACTCCGAGAGCCCTGCTGGCTGTCTTGGGCATCAGTGGCTATTTCCTCATG 900
QY 301 Asn 301
Db 901 AAT 903
RESULT 9
ID AAT59508
XX AAT59508 standard; cDNA; 1065 BP.
XX AAT59508;
XX AC
XX XX
DT 27-OCT-1997 (first entry)


```
Qy 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu 55
Db 171 ACATTTTCAAAATGTTCTACAAATGTATCCTACCAAGAACTACAAACCTAGTACCCTT 112
Qy 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
Db 111 GGAAGTACCAGCTGCACCTGTGTCTCAACATGGCAATGAGGCCACCAACAACATCACA 52
RESULT 12
ABA69132/C
ID ABA69132 standard; DNA; 233 BP.
XX
AC ABA69132;
XX
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #17437.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
PS Claim 4; SEQ ID NO 17437; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. NO.: 5.45e-16 Length: 233
Score: 311.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.41% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-2 (1-373) x ABA69132 (1-233)

Qy 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
Db 231 TCTGGGTTCATGAGTCTTGACAAACAGGACTGCTACCCAGAGATTACCTACCCAGGGA 172
```

```
Qy 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu 55
Db 171 ACATTTTCAAAATGTTCTACAAATGTATCCTACCAAGAACTACAAACCTAGTACCCTT 112
Qy 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
Db 111 GGAAGTACCAGCTGCACCTGTGTCTCAACATGGCAATGAGGCCACCAACAACATCACA 52
RESULT 13
ABA36064/C
ID ABA36064 standard; DNA; 233 BP.
XX
AC ABA36064;
XX
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #14530 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
PS Claim 4; SEQ ID No 14530; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. NO.: 5.45e-16 Length: 233
Score: 311.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.41% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-2 (1-373) x ABA36064 (1-233)
```

```
QY      16  SerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGly  35
      |||
Db      231 TCTGGTTCATGAGCTCTGACAAACAGGTACTGCTACCCAGAGTTACCTACCCAGGGA 172

QY      36  ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeu  55
      |||
Db      171 ACATTTTCAATGTTTCTACAAATGTATCCTACCAAGAAACTACAAACACCTAGTACCCCT 112

QY      56  GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr  75
      |||
Db      111 GGAAGTACCACCGCTGACCCCTGTCTCAACATGCAATGAGGCCACCAACAAATCACA 52

RESULT 14
AAK17442/C
ID      AAK17442 standard; DNA; 233 BP.
XX
AC      AAK17442;
XX
DT      05-NOV-2001 (first entry)
XX
DE      Human brain expressed single exon probe SEQ ID NO: 17433.
XX
KW      Human; brain expressed exon; gene expression analysis; probe;
KW      microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW      epilepsy; cancer; ss.
XX
OS      Homo sapiens.
XX
PN      WO200157275-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US00667.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-483446/52.
XX
PT      Single exon nucleic acid probes for analyzing gene expression in human
PT      brains -
XX
PS      Example 4; SEQ ID NO: 17433; 650pp + Sequence Listing; English.
XX
CC      The present invention provides a number of single exon nucleic acid
CC      probes which are derived from genomic sequences expressed in the human
CC      brain. They can be used to measure gene expression in brain cell samples,
CC      which may enable the diagnosis and improved treatment of nervous system
CC      diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC      epilepsy and cancers. The present sequence is one of the probes of the
CC      invention.
XX
SQ      Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. No.:      5.45e-16      Length:      233
Score:          311.00      Matches:      60
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    16.41%      Indels:      0
DB:            22      Gaps:      0

US-09-836-602-2 (1-373) x AAK17442 (1-233)
```

```
QY      16  SerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGly  35
      |||
Db      231 TCTGGTTCATGAGCTCTGACAAACAGGTACTGCTACCCAGAGTTACCTACCCAGGGA 172

QY      36  ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeu  55
      |||
Db      171 ACATTTTCAATGTTTCTACAAATGTATCCTACCAAGAAACTACAAACACCTAGTACCCCT 112

QY      56  GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr  75
      |||
Db      111 GGAAGTACCACCGCTGACCCCTGTCTCAACATGCAATGAGGCCACCAACAAATCACA 52

RESULT 15
AAK43242/C
ID      AAK43242 standard; DNA; 233 BP.
XX
AC      AAK43242;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human bone marrow expressed single exon probe SEQ ID NO: 17799.
XX
KW      Human; bone marrow expressed exon; gene expression analysis; probe;
KW      microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS      Homo sapiens.
XX
PN      WO200157276-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US00668.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488900/53.
XX
PT      Human genome-derived single exon nucleic acid probes useful for
PT      analyzing gene expression in human bone marrow -
XX
PS      Example 4; SEQ ID NO: 17799; 658pp + Sequence Listing; English.
XX
CC      The present invention provides a number of single exon nucleic acid
CC      probes which are derived from genomic sequences expressed in the human
CC      bone marrow. They can be used to measure gene expression in bone marrow
CC      samples, which may enable the improved diagnosis and treatment of cancers
CC      such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC      the probes of the invention.
XX
SQ      Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. No.:      5.45e-16      Length:      233
Score:          311.00      Matches:      60
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    16.41%      Indels:      0
DB:            22      Gaps:      0

US-09-836-602-2 (1-373) x AAK43242 (1-233)
```

```
QY      16  SerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGly  35
```

Db 231 TCTGGGTTTCATGAGTCTTGACAAACGGTACTGTACCCACAGAGTTACCTACCCAGGGA 172
QY 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu 55
Db 171 ACATTTTCAAATGTTTCTACAAATGTATCTACCAAGAAACTACAACACCTAGTACCCCT 112
QY 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
Db 111 GGAAAGTACCAGCCTGCACCCCTGTGTCTCAACATGGCAATGAGGCCACACAACATCACA 52

Search completed: October 30, 2002, 08:25:37
Job time : 178.922 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:08:08 ; Search time 1912.1 Seconds
(without alignments)
4082.221 Million cell updates/sec

Title: US-09-836-602-2
Perfect score: 1895
Sequence: 1 MPRGWTALCLLSLLPSGFMS.....QATSRNGHSARQHVVADTEL 373

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n_model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09836602/runat_29102002_101144_3785/app_query.fasta_1.1429
-DB=GenEmbl -FAST=Fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602@cgn.1.1.3800_runat_29102002_101144_3785 -NCPU=6 -TCPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1895	100.0	2463	9	S53910	S53910 CD34-glycop
2	1891	99.8	2615	9	HUMCD34HS	M81104 Human CD34
3	1827	96.4	2657	6	AX333753	AX333753 Sequence
4	1827	96.4	2657	9	S53911	S53911 CD34-glycop
5	1272	67.1	2956	4	CFU49457	U49457 Canis faml
6	1212	64.0	2690	4	AF461503	AF461503 Sus scrof
7	1173.5	61.9	1149	10	S69299	S69299 CD34-cell s
8	1173.5	61.9	1260	10	S69293	S69293 CD34-stem c
9	1173.5	61.9	2398	10	BC006607	BC006607 Mus muscu
10	1061.5	56.0	2453	4	AB021662	AB021662 Bos tauru
11	1019.5	53.8	2420	4	AF259378	AF259378 Capra hir
12	980	51.7	581	9	AF202879	AF202879 Homo sapi
13	904	47.7	1065	6	AR078454	AR078454 Sequence
14	870.5	45.9	978	10	S69301	S69301 CD34-cell s
15	630.5	33.3	160771	9	H888L2	AL035091 Human DNA
16	630.5	33.3	182742	2	AC084393	AC084393 Homo sapi
17	617	32.6	185162	2	AL365178	AL365178 Homo sapi
18	617	32.6	212535	2	AL356275	AL356275 Homo sapi
19	554	28.2	98347	2	AL513203	AL513203 Mus muscu
20	554	28.2	168333	2	AL513470	AL513470 Mus muscu
21	409	21.6	254	9	HUMCD34S3	M81940 Human CD34
22	325	17.2	1580	11	G06650	G06650 human STS W
23	321	16.9	1386	9	HUMCD34S8	M81945 Human CD34
24	311	16.4	183	9	HUMCD34S2	M81939 Human CD34
25	307	16.2	1330	10	S69295S2	S69302 stem cell a
26	275	14.5	165	9	HUMCD34S7	M81944 Human CD34
27	260	13.7	157	9	HUMCD34S5	M81942 Human CD34
28	233	12.3	235	6	AX247364	AX247364 Sequence
29	188	9.9	3124	5	MMTH4D1	Y13978 Gallus gall
30	168.5	8.9	12412	3	CELH43E16	AL591180 Zebrafish
31	167	8.8	80272	5	AL591180	AF109393 Rattus no
32	166.5	8.8	5509	10	AF109393	AL591175 Zebrafish
33	166	8.8	89232	5	AL591175	U11583 Saccharomyc
34	162.5	8.6	55069	8	YSGH9196	U20824 Equine herp
35	162.5	8.6	184427	14	EHVU20824	AL669823 Mus muscu
36	160.5	8.5	253273	2	AL669823	AB020726 Rattus no
37	159.5	8.4	1731	10	AB020726	AC019041 Homo sapi
38	158.5	8.4	183358	9	AC019041	L02115 Frog Integu
39	156.5	8.3	2161	5	XELFIMC1X	AC007263 Homo sapi
40	155.5	8.2	167390	9	AC007263	AC074046 Mus muscu
41	155.5	8.2	182534	10	AC074046	Z71659 S.cerevisia
42	154	8.1	3053	8	SCVNR044W	M60590 S.cerevisia
43	154	8.1	3137	8	YSCAAGLCS	AL357093 Human chr
44	153.5	8.1	167254	9	CNS05TDS	AC104020 Homo sapi
45	153.5	8.1	172789	2	AC104020	

ALIGNMENTS

RESULT 1

S53910
LOCUS S53910 2463 bp mRNA linear PRI 08-MAY-1993
DEFINITION CD34-glycoprotein expressed in lymphohematopoietic progenitor cells
(alternatively spliced) [human, UT7 cells, mRNA, 2463 nt].
ACCESSION S53910
VERSION S53910.1 GI:264766
KEYWORDS human UT7 cells.

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Nakamura,Y., Komano,H. and Nakauchi,H.
TITLE Two alternative forms of cDNA encoding CD34
JOURNAL Exp. Hematol. 21 (2), 236-242 (1993)
MEDLINE 93146100
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbs3 124007] from the original journal article.
This sequence comes from Fig. 1A.
FEATURES
source
1..2463
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene
1..2463
CDS
91..1248
/gene="CD34"
/translation="MLVRRGARAGPRGWTALCLLSLPSGFMSLDNNGTATPELP
TQTFESNVSTNVSYQETTPSTLGSSTSLHPVSOHGNEATTNITETTVKSTVITSV
YGNWNSVQSOTSIVSTVFTTPANVSTPETTLAPSLSPGNVSDLTSTSTSLATSPTRP
YTSSPLSDIKAEIKCSGIREVKLTQGLCNEQNTSSCAEFKKDRGEGEARLVLCGEE
QADAGAQCISLLAQSEVRPQCLLVLANRTEISIKLQMKRKHGSDLKGLIDFT
EQDVASHQSYQSKTLTALVTSALLAVLGITGYFLMNRNRSWPTGRRLGDPYYTENG
GGGYSGGGPTSPGACGKASVNRGAQNGTQATSRNGHSARQHVVADTEL"
BASE COUNT 589 a 709 c 580 g 585 t
ORIGIN
Alignment Scores:
Pred. No.: 4,45e-115 Length: 2463
Score: 1895.00 Matches: 373
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-836-602-2 (1-373) x 853910 (1-2463)
QY 1 MetProArgGlyTTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCCGGGGGCTGGACCGCCTTGCTGTGAGTTGTGCTTCTGGGTTTCATGAGT 186
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 246
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTCAAAATGATCTCTACCAAGAACTACACACCTAGTACCTTGGAACTACGAGCTG 306
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValIlys 80
Db 307 CACCCTGTGCTCAACATGGCAATGAGCCCAACAACATCACAGAAACACAGTCAAA 366
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATAACCTCAGTTTATGAAACACAAACTCTCTGTCTCCAGTCA 426
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTGAATCAGCACAGTGTTCACCAACCCAGCCAACTTCAACTCCAGAGACA 486
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 487 ACCTTGAAGCCCTGCTGCTACCTGGAAATGTTTCAGACCTTTCACCACTAGCAGTACG 546

```

```

QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
Db 547 CTTGCAACATCTCCACTAAACCTATACATCATCTCTCTATCCTAAGTGACATCAAG 606
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 607 GCAGAAATCAAAATGTTCCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCTGCCTGGAG 666
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 667 CAAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAGACAGGGAGGGGCTGGCCCGA 726
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 727 GTGCTGTGTTGGGAGGAGCAGGCTGATGCTCTGGGCCCCAGGTATGCTCCTGCTC 786
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 787 CTTGCCAGCTGTGAGGTGAGGCCCTCAGTGTCTACTGTGTGCTTTGGCCACAGAACAGAA 846
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 847 ATTTCCAGCAAACTCCAACTTATGAAAGACCAATCTGACCTGAAAGAGCTGGGATC 906
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 907 CTAGATTTCAGTGAGCAAGATGTTGCAAGCCACAGAGCTATTTCCCAAAAGACCTGATT 966
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 967 GCACGTGTCACCTCGGAGGCCCTGCTGCTCTGTTGGCATCAGTGGCTATTTCTCTGATG 1026
QY 301 AsnArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrThrGlu 320
Db 1027 AATGCCCGAGCTGGAGCCCCACAGAGAAAGGCTGGCGAAGACCTTATACACGGAA 1086
QY 321 AsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
Db 1087 AACGGTGGAGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTCAGGCTCAGGGAAG 1146
QY 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
Db 1147 GCCAGTGTGAACCGAGGGGCTCAGAAACGGGACCGCCAGCCACCTCCAGAAACGGC 1206
QY 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 1207 CATTCAACAGACAAACACAGCTGGTGGCTGATACCGAATTG 1245
RESULT 2
HUMCD34HS HUMCD34HS 2615 bp mRNA linear PRI 02-NOV-1993
LOCUS Human CD34 mRNA, complete cds.
DEFINITION M81104 X60172
ACCESSION M81104.1 GI:180108
VERSION M81104.1
KEYWORDS CD34; hematopoietic stem cell surface antigen; sialomucin.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Simmons,D.L., Satterthwaite,A.B., Tenen,D.G. and Seed,B.
TITLE Molecular cloning of a cDNA encoding CD34, a sialomucin of human
hematopoietic stem cells
JOURNAL J Immunol. 148, 267-271 (1992)
MEDLINE 92091783
FEATURES
source
1..2615
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
294..1415
/codon_start=1
/product="CD34"
/protein_id="AAA03181.1"
CDS

```

/db_xref="GI:180109"
/translation="MPRGWTAALCLLSLPSGEMSLDNNNGTATPELPTQGFSSNV
SYEETTPSLGSLHPVSHOHNENATNTITVKFTSTSVTSVITNSVQSQ
SVLSTVFTTPANVSTETTLKPSLSPCNVSLSTSTSLATSPKPYTSSPILSDIK
AEIKCSGIRVXLTKGICLQONTSSCAEFKDRGRLGVLCEQADADACAGVCS
LLAQSEVRPOCLLLVLANTEISSQLMKKHOSDLKLGILDFTEQDVASHOSYSQ
KTLLALVTGALLAVLGITGYFYLMNRSSPSPTGERLGEDPYTENGSGGQYSSGPGTS
PEAQGRASVYRQAKNGTQATSRNGHSARQHVAVDEL"

BASE COUNT 616 a 763 c 627 g 609 t
ORIGIN

Alignment Scores:
Pred. No.: 8,79e-115 Length: 2615
Score: 1891.00 Matches: 372
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.73% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x HUMCD34HS (1-2615)

Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 294 ATGCCGGGGCTGGACCGCTTGTCTGCTGAGTTTGTCTGCTTCTGGGTTTCATGAGT 353
Qy 21 LeuAspAsnAsnGlyThrAlaLeuProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 354 CTTGACAAACAGCGTACTACCCAGAGTTACTACCCAGGGAACATTTTCAAGTGT 413
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 414 TCTACAAATGTATCTACCAAGAAATACAAACACTAGTACCTTGGAAAGTACCAGCGTG 473
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 474 CACCCTGTCTCAACATGGCAATGAGCCCAACAACATACAGAAACAGACAGTCAAA 533
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 534 TTCACATCTACTCTGTGTATACCTCAGTTATGATGAAACAACTCTTCTGTCAGTCA 593
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 594 CAGACCTCTGTAATCAGCACAGTGTTCACACCCAGCCAGCAAGTTTCAACTCCAGAGACA 653
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 654 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCAGTACG 713
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 714 CTTGCAACATCTCCCATTAACCTATACATCATCTCTCTATCTCTAAGTGACATCAAG 773
Qy 161 AlaGluLysCysSerGlyIleAArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 774 GCAGAAATCAAATGTTTCAGGCATCAGAGAACTGAAATGACTCAGGCGCATCTGCGTGGAG 833
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 834 CAAATTAAGACCTCAGCTGTGCGGAGTTTAAAGAGGACAGAGGAGGCGCTGGCCCGA 893
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 894 GTGCTGTGGGAGGAGAGCGCTGATGCTACTGTGCTGCTTGGCCCAACAGACAGAA 1013
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 954 CTTGCCAGTCTGAGGTGAGCGCTCAGTGTCTACTGTGCTTGGCCCAACAGACAGAA 1013
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysGlyIle 260
Db 1014 ATTTCCGCAAACTCCAACCTTATGAAAAGGACCAACCTGACCTGAAAGAGCTGGGATC 1073

Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1074 CTAGATTTCTACTGACGAAGATGTTGCAAGCCACAGAGCTATTTCCAAAAGACCTGATT 1133
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 1134 GCATGTGTACCTCGGAGCCCTGCTGCTGTCTGTGGCATCACTGGCTATTTCTCTGATG 1193
Qy 301 AsnArgArgSerTyrProThrGlyGluArgLeuGlyGluAspProTyrTyrThrGlu 320
Db 1194 ATCCCGCAGCTGGAGCCGCCACAGAGAGAGAGCTGGCGGAGAGACCTTATACAGGAA 1253
Qy 321 AsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
Db 1254 AACGGTGGAGGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTGAGGCTCAGGAAAG 1313
Qy 341 AlaSerValAsnArgGlyAlaGlnGlnAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
Db 1314 GCCAGTGTGAACCGAGGGGCTCAGAAAACCGGAGCCGCGCCAGGCCACCTCCAGAAACGGC 1373
Qy 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 1374 CATTACAGACACACACAGCTGGTGGCTGATACCGAATG 1412
RESULT 3
AX333753
LOCUS AX333753 2657 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 4262 from Patent WO0194629.
ACCESSION AX333753
VERSION AX333753.1 GI:18124472
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 4262 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
source 1..2657
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 624 a 790 c 609 g 634 t
ORIGIN
Alignment Scores:
Pred. No.: 1.38e-110 Length: 2657
Score: 1827.00 Matches: 372
Percent Similarity: 84.93% Conservative: 0
Best Local Similarity: 84.93% Mismatches: 1
Query Match: 96.41% Indels: 66
DB: 6 Gaps: 1
US-09-836-602-2 (1-373) x AX333753 (1-2657)
Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCCGGGGCTGGACCGCTTGTCTGCTGAGTTTGTCTGCTTCTGGGTTTCATGAGT 186
Qy 21 LeuAspAsnAsnGlyThrAlaLeuProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACAAACAGCGTACTGCTTACCCAGAGTTACTACCCAGGGAACATTTTCAAGTGT 246
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTCAAAATGTATCTTCAACAGAACTACAAACCTAGTACCTTGGAAAGTACCAGCGCTG 306
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80

```
Db 307 CACCCTGTGTCTCAACATGGCAATGAGGCCACACAAACATCATCAGAAAACGACAGTCAAA 366
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACCTCTCTGTGCCAGTCA 426
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTATCAGCACAGTGTTCACCACCCAGCCCAACGTTTCAACTCCAGAGACA 486
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 487 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTTCAGACCTTTCAACCACTAGCACTAGC 546
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 547 CTTGGACATCTCCCACTAAACCTTATACATCATCTTCTCTATCTTAAGTAGACATCAAG 606
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 607 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTGACTCAGGGCATCTCCCTGGAG 666
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 667 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAAGACAGAGGGCGCTGGCCCGA 726
Qy 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 727 GTCTGTGTGGGAGGAGCAGGCTGATGTGCTGTGGGGCCAGGTATGCTCCCTGCCT 786
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 787 CTTGCCAGTCTCAGGTGAGGCCTCAGTGTCTACTGCTGTGTCCTGGCCACACAGAA 846
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 847 ATTTCAGCAAACTCCAACCTATGAAAGACACCAATCTGACCTGAAAGAGCTGGGGATC 906
Qy 261 LeuAspPheThrGlnGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 907 CTAGATTCTACTGACCAAGATGTTGCACGCCACAGAGCTATTCCCAAGAACCCCTGATT 966
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 967 GCACCTGGTCACTCGGAGCCCTGCTGGCTGTCTTGGGCATCATCTGGCTATTTCCTGATG 1026
Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeu----- 312
Db 1027 AATCGCCGACGCTGGAGCCCCACAGAGAAAGGCT-GGAGCTGGAAACCCCTGACCACTCTT 1085
Qy 312 ----- 312
Db 1086 CAGGAAGAAGGAGTCTGCACATGACGCTGCACCCCTCCCTCCGATCCTCTCCACCTC 1145
Qy 312 ----- 312
Db 1146 CCCTCCCGCTTCTCCACCCCTGCCCCACCTTCTCTGTTGGGCCCTCTCCCATCCAGTG 1205
Qy 313 -----GlyGluAsp 315
Db 1206 TCTCAGAGCCCTGCTTACCAGATAAATGCTACTTTATTTATACATGCTCTAGGGCGAAGAC 1265
Qy 316 ProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerPro 335
Db 1266 CCTATTATACGGAAACCGTGGAGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTT 1325
Qy 336 GluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGlnAsnGlyThrGlyGlnAla 355
Db 1326 GAGGCTCAGGAAAGCCAGTGTGAACCGAGGGGCTCAGGAAACCGGGACCGCCAGGCC 1385
Qy 356 ThrSerArgAsnGlyHisSerAlaArgGlnHisValAlaAspThrGluLeu 373
Db 1386 ACCTCCAGAAACGGCCATTTCAGCAGACAAACACGCTGGTGGCTGATACCGAATTG 1439

RESULT 4
LOCUS S53911 2657 bp mRNA linear PRI 08-MAY-1993
DEFINITION CD34-glycoprotein expressed in lymphohematopoietic progenitor cells
(alternatively spliced, truncated form) [human, UT7, mRNA, 2657
nt].
ACCESSION S53911
VERSION S53911.1 GI:264768
KEYWORDS human UT7.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2657)
AUTHORS Nakamura,Y., Komano,H. and Nakauchi,H.
TITLE Two alternative forms of cDNA encoding CD34
JOURNAL Exp. Hematol. 21 (2), 236-242 (1993)
MEDLINE 93146100
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 124009] from the original journal article.
This sequence comes from Fig. 1AB.
FEATURES
source
1..2657
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene
1..2657
/gene="CD34"
CDS
91..1077
/gene="CD34"
/note="glycoprotein expressed in lymphohematopoietic
progenitor cells; This sequence comes from Fig. 1AB"
/codon_start=1
/product="CD34"
/protein_id="AAB25223.1"
/db_xref="GI:264769"
/translation="MLVRRGARAGPRMGWTAICLLSLPSGFMSLONNGTATPELP
TQGTFSNVSTNVSYOETTPSTLGSTSLHPVSOHQNEATTNITETVKTSTVTSY
YGNNTSSVSQSTSVISTVETTPANYSPTETLKPSLSPGNVSDLTSTTSLATSPTKP
YTSSPILSDIAKEIKCSGIREVKLTQGLCEQNTSSCAEFKKDRGGRLARVLGEE
QADAGAGQVCSLLAQSEVRPCILLVLANRTEISSKLQMKKHQSDLUKRLIDFT
EDVASHQSYSQKTLIALVTGALLAVLGITGYFLMNRSSWSPSTGERLELEP"
BASE COUNT 624 a 790 c 609 g 634 t
ORIGIN
Alignment Scores:
Pred. No.: 1.38e-110 Length: 2657
Score: 1827.00 Matches: 372
Percent Similarity: 84.93% Conservative: 0
Best Local Similarity: 84.93% Mismatches: 1
Query Match: 96.41% Indels: 66
DB: 9 Gaps: 1
US-09-836-602-2 (1-373) x S53911 (1-2657)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCCCGGGCTGGACCGGCTTCTGCTGAGTTTGTGCTTCTGGGTTTCATGAGT 186
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACAAACACGGTACTGCTACCCAGAGTTACCTACCCAGGAAACATTTCAATGTT 246
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTACAAATGATCTCTACCAAGAAACTACAAACACCTAGTACCTTGGAAAGTACCACTG 306
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrAsnIleThrCluThrValLys 80
Db 307 CACCCTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA 366
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
```


Db	367	TTCACATCTACCTCTGTGATACCTCAGTTTATGGAAACACAAACCTCTTCTGTCCAGTCA	426
QY	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	427	CAGACCTCTGTATACAGCACAGTGTACACACCCAGCCAACTTCACTCCAGAGACA	486
QY	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer	140
Db	487	ACCTTGAAGCCTAGCTGTACCTGGAATGTTCAGACCTTCAACACCACTAGCAGTAGC	546
QY	141	LeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIleLys	160
Db	547	CTTGCAACATCTCCCACTAAACCTATACATCTCTCCATCTCTAAGTGACATCAAG	606
QY	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	607	GCAGAAATCAATGTTCAGGCATCAGAGAAGTGAATTGACTCAGGGCATCTGCCTGGAG	666
QY	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	667	CAAAATAAGACCTCCACAGCTGTGGGAGTTTAAGAGACAGAGGGGAGGGCTGGCCCGA	726
QY	201	ValLeuCysGlyCluCluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	727	GTGCTGTGGGAGGAGCAGGCTGATGTGCTGTGGGGCCCGAGGTATGCTCCCTGCTC	786
QY	221	LeuAlaGlnSerCluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
Db	787	CTTGCCCACTCTGAGGTGAGGCTCACTGTCTACTCTGTCTGTGGCCCAACAGACAGAA	846
QY	241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle	260
Db	847	ATTTCAGCAAACTCCAACTTATGAAAAAGCACCACCATCTGCACCTGAAAAAGCTGGGGATC	906
QY	261	LeuAspPheThrCluGlnAspValAlaSerHisGlnSerLysThrGlnLysThrLeuIle	280
Db	907	CTAGATTTCAGTGAAGAAGTGTGAAGCCACAGAGCTATTCCCAAAAGACCCCTGATT	966
QY	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyThrPheLeuMet	300
Db	967	GCACCTGGTCACCTCGGAGCCCTCGTGGCTGTCTGGGCATCTACTGGCTATTTCCTGATG	1026
QY	301	AsnArgArgSerTrpSerProThrGlyGluArgLeu	312
Db	1027	AATCGCCGAGCTGGAGCCCCACAGGAGAAAGGCT-GGAGCTGGAACCCCTGACCACTCTT	1085
QY	312	-----	312
Db	1086	CAGGAAGAAAGAGTCTGCACATGCAGCTGCACCCCTCCCTCCGATCCTTCTCCACCTC	1145
QY	312	-----	312
Db	1146	CCCTCCCTCTCTCCACCCCTGCCCCACTTCCTGTTGGCCCTCTCCCATCCAGTG	1205
QY	313	-----GlyGluAsp	315
Db	1206	TCTCACAGCCCTGCTTACCAGATAATGCTACTTTATTATATACACTCTAGGCGCAAGAC	1265
QY	316	ProThrThrThrCluAsnGlyGlyGlyGlnGlyTyThrSerSerGlyProGlyThrSerPro	335
Db	1266	CCTTATTACCGGAAACCGTGGAGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTC	1325
QY	336	GluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAla	355
Db	1326	GAGGCTCAGGAAAGCCAGTGTGAACCGAGGGGCTCAGGAAACCGGACCGCCAGGCC	1385
QY	356	ThrSerArgAsnGlyHisSerAlaArgGlnHisValValAlaAspThrGluLeu	373
Db	1386	ACCTCCAGAAACGGCAATTTCAGCAACAACACGCTGGTGGCTGATACCGGAATTG	1439
RESULT	5		
CFU49457			
LOCUS			

2956 bp mRNA linear MAM 21-JAN-1997

DEFINITION	Canis familiaris hematopoietic progenitor cell marker CD34 mRNA, complete cds.			
ACCESSION	U49457			
VERSION	U49457.1			
KEYWORDS	GI:1224105			
SOURCE	dog.			
ORGANISM	Canis familiaris			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
AUTHORS	1 (bases 1 to 2956) McSweeney,P.A., Rouleau,K.A., Storb,R., Bolles,L., Wallace,P.M., Beauchamp,M., Krizanac-Bengez,L., Moore,P., Sale,G., Sandmaler,B., de Revel,T., Appelbaum,F.R. and Nash,R.A.			
TITLE	Canine CD34: cloning of the cDNA and evaluation of an antiserum to recombinant protein			
JOURNAL	Blood 88 (6), 1992-2003 (1996)			
MEDLINE	96420219			
AUTHORS	2 (bases 1 to 2956) McSweeney,P.A.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-FEB-1996) P.A. McSweeney, Transplant Biology, FHCR, 1124 Columbia St, Seattle, WA 98104, USA			
FEATURES	Location/Qualifiers			
source	1..2956 /organism="Canis familiaris" /db_xref="taxon:9615" /tissue_type="ML2 myelomonocytic leukemia"			
gene	1..2956 /gene="CD34"			
CDS	260..1429 /gene="CD34" /note="type 1 transmembrane protein" /codon_start=1 /product="hematopoietic progenitor cell marker CD34" /protein_id="AAB41055.1" /db_xref="GI:1224106" /translation="MLAGRARAGGLPRGWTALCLLSLPFGFTNVTETVPTPTSTEINSAVSENTSKRAITLTPTSGTTLTVSVQSDSGTTATISETHVHTSTSEITLTPTMNSVQSOTSLAITVSTFTNFSTSVTLEPSLLPGNSDPYNSVLVTSPTETVYLSPTPSRNDTPSTIKGEIKCSGVKVKLNQIGCLENETSSCEDFKDNEELTKQVICREPAEAGAVCSLLLAQSEVRPHCLLVLANKEITFSKLQLLKRHQSDULKIGIRDFTDDVGHOSHYSKRLLIALVTGILLAVLTGTGYFLMNRNRSPTGERIGEDPYITENGGQGYSSGPGVSPQAQSKASVNRGPGQENGTSRNGHSARQHMVADTEL"			
BASE COUNT	625 a	864 c	815 g	652 t
ORIGIN				
Alignment Scores:				
Pred. No.:	3,28e-74	Length:	2956	
Score:	1272.00	Matches:	265	
Percent Similarity:	77.04%	Conservative:	27	
Best Local Similarity:	69.92%	Mismatches:	79	
Query Match:	67.12%	Indels:	8	
DB:	4	Gaps:	2	
US-09-836-602-2 (1-373) x CFU49457 (1-2956)				
QY	1	MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20	
Db	296	CTGCCCGGGCTGGACCGGCTCTGCTGCTCAGTCTGCTGCTTGGGTTCACAAAC	355	
QY	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40	
Db	356	ACAGAAACCGTGATTACTCTTACCACAGTGCACACAAATAATGTGAGCTGTT	415	
QY	41	SerThrAsnValSerThrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu	60	
Db	416	TCTGAGAATACATCCAAACGGGAAGCCATCACACTACTCTTCTGGAACACTACCCCTG	475	
QY	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGlnThrThrValLys	80	
Db	476	TACTCTCTCTCAAGACAGACAGCTGGGACCACAGCAACCATCTCAGAGACTCAGTCCAT	535	
QY	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100	

Db 536 GTACATATACCTCTGAGATCACCTTAACGCGCTGGGACCATGAACCTCTTCTTTCAGTCG 595
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 596 CAGACCTCTTTAGCTATACAGGTATCTTTTACCCCAACCAACTTTTCACTTCAAGTGTG 655
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 656 ACCTTGGAGCCAGCTGCTACCTGGAATGTTGCGGATCCCCCTACACAGCACCGAGC 715
Qy 141 LeuAlaThrSerProThrLysProThrThrSerSerPro----- 154
Db 716 CTTGTGACATCCCCACGGAATATTATACATCACTTCTCTACCCCAAGTAGAAATGAC 775
Qy 155 IleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThr 174
Db 776 ACCCAAGTACCATCAAGGGAAATCAATGTTCCGGAGTCAAGAAGTGAATTTGAAC 835
Qy 175 GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArg 194
Db 836 CAAGGTATCTGCCTAGAGCTAATAGACCTCCACCTGCTGAGGACTTTAAGAAAGATAAC 895
Qy 195 GlyGluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAla 214
Db 896 GAAGAAAAAAGTACCAAGTCTGTGT-----GAGAAAGAGGCCAGCTGAGGCTGGGGCC 949
Qy 215 GlnValCysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuVal 234
Db 950 GGGGTGTCTCCCTGCTCTGCGCCAGTCTGAGGTGAGGCTTCACCTGCTGCTGTGTC 1009
Qy 235 LeuAlaAsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAsp 254
Db 1010 TTGGCCCAACAAAACAGAACTTTTCAGTAACTCCAACTCTCTGAGAAAGCACCACTGAC 1069
Qy 255 LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyr 274
Db 1070 CTGAAAAAGCTGGGGATCCGACATTCATGAAACAGATGTTGGGAGCCACAGAGCTAT 1129
Qy 275 SerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIle 294
Db 1130 TCCGCCAAGACCTGATGTCATGCTGCTACCTCAGGAGTCTGCTGCTGCTGTGGGCCACC 1189
Qy 295 ThrGlyTyrPheLeuMetAsnArgSerTrpSerProThrGlyGluArgLeuGlyGlu 314
Db 1190 ACTGTTACTTCTGATGACCGCGCAGTGGAGCCCTACAGGAGAAAGCTGGCGGAA 1249
Qy 315 AspProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSer 334
Db 1250 GACCCTTATTACCGGAGAACGGTGGAGGCCAGGCTATAGTCAAGGCCCTGGGGTCTCC 1309
Qy 335 ProGluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGln 354
Db 1310 CTTGAGGCTCAGGGAAGCCAGTGTGAACCTGGGCCCTCAGGAGAACGGGACCGGCCAG 1369
Qy 355 AlaThrSerArgAsnGlyHisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 1370 GCCAGTCCAGAAAGGGCCATTTCAGCAGACACACACATGGTGGCTGTATACAGAAATTG 1426
RESULT 6
AF461503
LOCUS AF461503 2690 bp mRNA linear MAM 24-JAN-2002
DEFINITION Sus scrofa CD34 antigen mRNA, complete cds.
ACCESSION AF461503
VERSION AF461503.1 GI:18308137
KEYWORDS .
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 2690)
AUTHORS Sun, J., and Butler, J.E.
TITLE Cloning, sequencing and expression of swine CD34 gene

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2690)
AUTHORS Sun, J., Wang, J., and Butler, J.E.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-2001) Microbiology, University of Iowa, 51 Newton Road, Iowa City, IA 52242, USA
FEATURES
source
1. .2690
/organism="Sus scrofa"
/db_xref="taxon:9823"
/tissue_type="mesenteric lymph nodes"
298. .1467
/note="cluster of differentiation antigen 34"
/codon_start=1
/product="CD34 antigen"
/protein_id="AA167838.1"
/db_xref="GI:18308138"
/translation="MPRGWTTCLLSLLPSGFTAVNSSTVASTLPTAAGSTPTGTPATA
TSAITLGPETVNSVQPTSLATVSSATINFTTSEVTLQPTFPNGVSDPLYNSTSPA
RSPNTPYSSPPTPGSHKGEVKCAQIKVEKLTOGICLERNETSGCEKFKDNGEKLMQ
ILCGEQEAERAGPGVCSLLLAQSEVKPHCLLLVLANTELSSKFLLEKHKQSELRMSI
QHFSKQDVRHQSYSRKTLTALVTSGILLAVLGTIGTGLLMNRRNSWSPTGRLGDPYI
TENGGGGYSSGPGASPEAQGKASVTRGAQENGTTQATSRNGHSARQPMVADTEL."
BASE COUNT 547 a 808 c 738 g 597 t
ORIGIN
Alignment Scores:
Pred. NO.: 2.44e-70 Length: 2690
Score: 1212.00 Matches: 255
Percent Similarity: 73.60% Conservativeness: 35
Best Local Similarity: 64.72% Mismatches: 78
Query Match: 63.96% Indels: 26
DB: 4 Gaps: 6
US-09-836-602-2 (1-373) x AF461503 (1-2690)
Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 298 ATGCCGCGGGCTGGACCCAGCTCTGCTTGTGCTGAGTTTGTCTGCTCTGCGCTTCACAGCT 357
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuPro----- 32
Db 358 GTG---AACAGCTGACGCTGTTGCTTCCACCTTGCCAACTGGCGCTGCAACTCCCACC 414
Qy 33 -----ThrGlnGlyThr-----PheSerAsn 39
Db 415 GGGCCAGCTACTGCGAGGACAGCTATCACCGGTCAGCTATCTCAGACATATCTTCACT 474
Qy 40 ValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSer 59
Db 475 GTTCTCAAAATATATATCCACAGAGAA---ACCACACACAGATGCTTTCGAAGTCCAGC 531
Qy 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVal 79
Db 532 CTCCACACTGCTCTCTCAGGCGCAGTGGGACCCAGCCGAGCATCTCAGGCCCTACAGTT 591
Qy 80 LysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGln 99
Db 592 AATTTCATGTCTACCTCGCGGCTCACCTCTCCCGCAACCCGTCGAACCTCTCTGTCTCAG 651
Qy 100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
Db 652 CCTCAGACCTCT---CTAGCCACAGTGTCTCCGCCACCATCACTTTACAACTTCAGAG 708
Qy 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrSerThr 139
Db 709 GTGACCTTCAGCCAGCAGCAGTTCACAGAAATGTTTTCAGACCCCTCTACAAACAGTACC 768
Qy 140 SerLeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIle 159
Db 769 AGCCCTCGGAGATCCCCACACGCCCCCTACACATCATCTCTCTCTACCCAGGTAGCCAC 828

```
Qy 160 LysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Db 829 RAAGGGGAAGTCAATGTGCCAATCAAGAGAGTGAATGACCAAGGTATCGCTG 888
Qy 180 GluGlnAsnLysThrSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Db 889 GAGCGAAATGAGACCTCGCGCTGCGAGAGTGAATGAAGAGCAATCGAGAGAGTGTATG 948
Qy 200 ArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
Db 949 CAATTCCTGTGTGGCAGAGCAGCTGAGCGGGGCCAGG-----GTGTGCTCCTTG 1002
Qy 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThr 239
Db 1003 CTCCTTGCCCAATCTGAGTGAACCTCAGCTGCCTGCTGCTGTGCTTGGCCAACGGAACA 1062
Qy 240 GluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLysLysLeuGly 259
Db 1063 GAACCTTAGCAGCAAGTTCCTGCTTCTGGAAACGACCACTCTGAACCTGAGAGATGAGC 1122
Qy 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu 279
Db 1123 ATCCAACACTCTCGAACACAAGATGTAGGAGCCACAGAGCTACTCCGGAAGACCTTG 1182
Qy 280 IleAlaLeuValThrSerGlyAlaLeuAlaValLeuGlyIleThrGlyTyrPheLeu 299
Db 1183 ATTGCACCTGTCACCTCGGGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1242
Qy 300 MetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrThr 319
Db 1243 ATGAACCGCTCGAGTGGAGCCTACAGGAGAAAGCTGGGGAAGACCTTATTACAGC 1302
Qy 320 GluAsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGly 339
Db 1303 GAGACGGTGGAGGAGGCTATAGCTAGCGCCCTGGGGCCCTCCCTGAGGCTCAGGGA 1362
Qy 340 LysAlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsn 359
Db 1363 AAAGCCAGTGTCACTCAGAGGGCTCAGGAGAACGCGCCAGCCAGCCACCTCCAGAAC 1422
Qy 360 GlyHisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 1423 GGCCATTACAGAGGCAACCCATGTGTGGCAGATACTGAAC 1464

RESULT 7
S69299
LOCUS S69299 1149 bp mRNA linear ROD 03-JUN-1994
DEFINITION CD34-cell surface antigen (alternatively spliced, clone 6-2-2)
[imice, stromal cell line PA-6, mRNA Partial, 1149 nt].
ACCESSION S69299
VERSION S69299.1 GI:495713
KEYWORDS Mus sp. stromal cell line PA-6.
SOURCE Mus sp.
ORGANISM Mus sp.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1149)
TITLE Suda, J., Sudo, T., Ito, M., Ohno, N., Yamaguchi, Y. and Suda, T.
JOURNAL Two types of murine CD34 mRNA generated by alternative splicing
MEDLINE Blood 79 (9), 2288-2295 (1992)
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 145456] from the original journal article.
This sequence comes from Fig. 3.
FEATURES
source
1..1149
/organism="Mus sp."
/db_xref="taxon:10095"
gene
1..1149
/partial
/gene="CD34"
CDS
1..1149
/gene="CD34"
```

```
/note="cell surface antigen; This sequence comes from Fig.
3"
/codon_start=1
/protein_id="AAB22108.1"
/db_xref="gi:495714"
/translation="MOVHRTFRAGLLLPWRWVALCLMSLLHLLNLTSTATTSTQGIS
PSVTNSESVEINITSSIPGSTSHYLIYDSSKTTTPAISTVMVNTVTSGIPSGSTPH
TFSOPOTSPGILPTTSDSISTSEMTWKSSLPISVSDYSPNNSSFEMTSPTEPVAYT
SSASPAIKGEIKCSGIREVRLAIGLELSEASCEFEKKEGKEDLILQILCEKEAE
ADASVCSLLLAQSEVRPECLMLVLANSTELPSKLQLMKHXGLDLRLGLQSFNKQD
IGHSQSYSRKTALIALVTGVLALTGTGYELMNRNWSPTGERLGEDPYTENGQGO
GYSGPGASPETQKANVTRGAQENGTCQATSRNGHSARQHVAVDEL"
BASE COUNT 307 a 293 c 285 g 264 t
ORIGIN
Alignment Scores:
Pred. No.: 2.66e-68 Length: 1149
Score: 1173.50 Matches: 245
Percent Similarity: 75.86% Conservative: 41
Best Local Similarity: 64.99% Mismatches: 80
Query Match: 61.93% Indels: 11
DB: 10 Gaps: 4
US-09-836-602-2 (1-373) x S69299 (1-1149)
Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 37 CTGCCATGGCGTGGGTAGCTCTCGCTGATGAGTCTGCTG-----CAT 81
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 82 CTAATAACTTGCATCTCTACACGAGACTTCTACACAGGAATATCCCATCAGTT 141
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 142 CCTACCAATAGTCTGTTGAGAAATAATACATCATGATCCCTCGGAAGTACCAAC 201
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 202 TACTTGATCTATCAGACAGAGTAAAGACACACAGCCATCTCAGAGATATGTTCAAC 261
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 262 TTTACAGTTTACCTCTGGGATCCTTCAGGCTCTGGAACCTCCACACACTTTTCAACA 321
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 322 CAGACTTCCCCAACTGGCATACTACTTCTCAGACAGATATTTCCTCAGAGATG 381
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 140
Db 382 ACCTGGAAGTCCAGCTGCCATCTATAATGTTCTGATATTTCGCTAATATAGCAGC 441
Qy 141 Leu---AlaThrSerProThrLysProTyr-----ThrSerSerProIleLeu 156
Db 442 TTTGAGATGACATCAACCCAGCAGCATATGCTTACACATCATCTTCTGCTCG----- 495
Qy 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db 496 AGTGCCATTAAAGGAGAAATCAATGCTCTGGAATCCGAGAAGTGAAGTGGCCCGGT 555
Qy 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 556 ATCTGCTTGGAACTAAGTGAAGCACTAGTGTGAGGAGTTTTAAAGAAAGGAGGAA 615
Qy 197 GlyLeuAlaArgValLeuCysGlyGluGlnGlnAlaAspAlaAspAlaGlnVal 216
Db 616 GATCTAATTCAAATACTGTGTGAAAGAGGAGGCTGAGGCTGATGCTGCTGCTAGTGC 675
Qy 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db 676 TGCTCCCTCTTCTGACCCAGTCTGAGGTAGGCTAGGCTGAGTGTGCTGATGCTGTGGCC 735
```


QY 317 TrrTrrThrGluAsnGlyGlyGlyGlnGlyThrSerSerGlyProGlyThrSerProGlu 336
 |||||
 Db 999 TATTACAGGAATGGTGGAGGCGAGGCTATAGCTCAGGACCTGGGGCCCTCCCTGAG 1058
 |||||
 QY 337 AlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGlnAsnGlyThrGlyGlnAlaThr 356
 |||||
 Db 1059 ACTCAGGAAAGGCCAATGTGACCCGAGGGCTCAGGAGAACGGACCGGCCCACT 1118
 |||||
 QY 357 SerArgAsnGlyHisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
 |||||
 Db 1119 TCCAGAACGGCCATTTCAGCAAGACACATGTGTGGTGTGACACAGAAGCTG 1169
 |||||

RESULT 9
 BC006607 2398 bp mRNA linear ROD 12-JUL-2001
 LOCUS BC006607 Mus musculus, Similar to CD34 antigen, clone MGC:11720
 DEFINITION IMAGE:3966337, mRNA, complete cds.
 ACCESSION BC006607
 VERSION BC006607.1 GI:13879273
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2398)
 Strausberg, R.
 Direct Submission
 Submitted (27-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: villalob@bcm.tmc.edu
 Villalón, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 17 Row: n Column: 8
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Similarity but not
 identity to protein.
 Location/Qualifiers
 1..2398
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="MGC:11720 IMAGE:3966337"
 /tissue_type="Mammary tumor. Metallothionien-TGF alpha
 model. 10 month old virgin mouse. Taken by biopsy."
 /clone_lib="NCI_CGAP_Mam1"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 62..1210
 /codon_start=1
 /product="Similar to CD34 antigen"
 /protein_id="AAH06607.1"
 /db_xref="GI:13879274"
 /translation="MQVHRTRAGLLLPWRVVALCLMSLLHLNLTTSATTTSTQGIS
 PSVPTNSVETNTIPSTIPGTSYLLIYQDSKSTTPAISFTWNTFTVTSIPSGSGTH
 TFSQPTSGILPTTSDSIEMTWKSLINVSVDYSPNNSSFTMPTPEYAT
 SSSAPSAIKGICSGIREVRLAQICILELSEASSCEFFKKEGEDIQILCEKEAE
 ADAGASVCSLLLAQSEVRPECLLMVLANSTELPSKLQLMKQSLRKLQTSFNKQD

IGSHQSVSRKTLIALVTSGVLLAILGTTGYFLMNRSSWSPTGERLGEDPYTENGSGQ
 GYSSGCGASPETOGKANVTRGAQENGCTGQATSRNGHSARHVVADTEL"
 BASE COUNT 623 a 599 c 551 g 625 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,95e-68 Length: 2398
 Score: 1173.50 Matches: 245
 Percent Similarity: 75.86% Conservative: 41
 Best Local Similarity: 64.99% Mismatches: 80
 Query Match: 61.93% Indels: 11
 DB: 10 Gaps: 4
 US-09-836-602-2 (1-373) x BC006607 (1-2398)
 QY 1 MetProArgGlyTrrThrAlaLeuCysLeuLeuLeuProSerGlyPheMetSer 20
 |||||
 Db 98 CTGCCATGGCGCTGGGTAGCTCTCTGCCTGAGTGTCTGCTG-----CAT 142
 |||||
 QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 |||||
 Db 143 CTAATAACTTGAAGTCTGCTACCCAGGAGACTTCTACACAGGAAATATCCCCATCAGTT 202
 |||||
 QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 |||||
 Db 203 CTTACCAATGAGTCTGTTGAGGAAATATCATCTAGCATCCCTGGGAAGTACCAGCCAC 262
 |||||
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValValys 80
 |||||
 Db 263 TACTTGATCTATCAGCAGCAGCAGTAAAGACACACAGCAGCATCTCAGAGACTATGTCAC 322
 |||||
 QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 |||||
 Db 323 TTTACAGTTACTCTGGGATCCCTTCAGGCTCTGGAACCTCCACACTTTTTCACACCA 382
 |||||
 QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 |||||
 Db 383 CAGACTCCCAACTGGCATACTGCTACTTCTCAGACAGTATTCTCCACTTCAGAGATG 442
 |||||
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 |||||
 Db 443 ACCTGGAAGTCCAGCCTCCCATCTATNAATGTTCTGATATTTCGCCTAATAATAGCAGC 502
 |||||
 QY 141 Leu---AlaThrSerProThrLysProTrr-----ThrSerSerSerProfileLeu 156
 |||||
 Db 503 TTTGAGATGACATCATCCACCCAGGACCATATGCTTACACATCATCTTCTGCTCCG----- 556
 |||||
 QY 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
 |||||
 Db 557 AGTGCCATTAAAGGGAGAAATCAATGCTCTGGAATCCGAGAAGTGAAGTGGCCCGAGGT 616
 |||||
 QY 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
 |||||
 Db 617 ATCTGCTGGAACTAAGTGAAGCATCTAGTTCTGTAGAGAGTTTAAAGAGAAAGAGGAGAA 676
 |||||
 QY 197 GlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
 |||||
 Db 677 GATCTAATTCAAACTACTGTGTGAAGAAGGAGGAGGCTGAGGCTGATGCTGGTGTAGTCTC 736
 |||||
 QY 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
 |||||
 Db 737 TGCTCCCTGCTTCTAGCCAGCTCTGAGGTAGGCTGAGTGTGCTGCTGCTGCTGCTGCTG 796
 |||||
 QY 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
 |||||
 Db 797 AATAGCAGAACTTCCCGAGCAAACTCCAGCTTATGGAAGAACAGCAATCTGACTTGAGA 856
 |||||
 QY 257 LysLeuGlyIleLeuAspPheThrCluGlnAspValAlaSerHisGlnSerTyrSerGln 276
 |||||
 Db 857 AAGCTGGGAGTCCAAAGCTTCAATAAACAAGATATCGGAGGCCACAGAGTATTCCCGA 916
 |||||
 QY 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
 |||||


```

Db 1067 CCTGAGGCTCCCGGAGCTCAGGAAAGGCCATTGTGAATCAGGGGCTCAGGAGAA 1126
Qy 351 GlyThrGluGlnAlaThrSerArgAsnGlyHisSerAlaArgGlnHisValValAlaAsp 370
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1127 GGCACCGCCAGGACCTCAGAAACGGCCATTACGAGAGAACACGTGTGGTGCCTGAT 1186
Qy 371 ThrGluLeu 373
|||||
Db 1187 ACCGAAGT 1195

RESULT 11
AF259378
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Capra hircus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Caprinae; Capra.
1 (bases 1 to 2420)
Leipprandt, J.R., Anson, D.S. and Jones, M.Z.
Cloning and sequence analysis of caprine CD34 cDNA
Unpublished
2 (bases 1 to 2420)
Leipprandt, J.R., Anson, D.S. and Jones, M.Z.
Direct Submission
Submitted (20-APR-2000) Biochemistry, Michigan State University,
Room 519 Biochemistry, East Lansing, MI 48824, USA
LOCATION/Qualifiers
FEATURES
source
1..2420
/organism="Capra hircus"
/db_xref="taxon:9925"
<1..1069
/feature="cell-surface antigen"
/feature="CD34"
/codon_start=2
/product="CD34"
/protein_id="AAP76145.1"
/db_xref="GI:8515726"
/translation="PSMEFTNTAVTGNATVSSNPGISSPPTTTSYWKMSNSTFRST
SLYNSDSGTTPAVTSPASTVNFSTASSETLPASGVNSVOPQSTLATASSTLIS
FATSEATLPSTPGNISDSLYNSTSPVTSSTINTSPFSPTONILKTEIKCSRKEVK
LAHSICLKNETSCEEFKNNEKLTRVLCQEQAEVCSLLIAQSEVYKQCLLLVLT
NRTSSKILRLKYSQDLREMGIDISEEDVSHQYSRKRLIALTVSGLLAILIT
TCYFLNRRSWSPAGERLGEDPYVENGSGGQYNSGSEASPEAQGKAIVNRGAQENGT
QOATSRNGHSARQVVDTEL"
BASE COUNT 565 a 700 c 590 g 565 t
ORIGIN

Alignment Scores:
Pred. No.: 8.38e-58 Length: 2420
Score: 1019.50 Matches: 216
Percent Similarity: 72.36% Conservative: 38
Best Local Similarity: 61.54% Mismatches: 88
Query Match: 53.80% Indels: 9
Db: 4 Gaps: 3

US-09-836-602-2 (1-373) x AF259378 (1-2420)

Qy 25 GlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnVal 44
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 35 GGTATGCGCAGAGTGCA---TCGAACCCAGGATATCTTCACCTGCCCTCCCACTACAG 91
Qy 45 SerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSer 64
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 92 TCCCTACTGGAAAGCAGTCAAGTACTTTTGAAGCACCAGCGCTCTACAAATGCTCT 151
Qy 65 Gln-----HisGlyAsnGluAlaThrThrAsnIleThrGluThrValLysPheThr 82
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 152 CAGGACAGCAATGGGACACGACGACCTCACCTCAGTTCACGGCTCTACAGTCAATTCACG 211

```

```

Qy 83 SerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValcInSerGlnThr 102
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 TCTGCCCTGAGACCCCTCCAGCCTCTGGAGCGGTGAAGTCTCTGTCTCAACACGAC 271
Qy 103 SerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlnThrThrLeu 122
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 272 TCTTTAGCACACGAGGATCTTCTACCTCATCAGCTTTGGGACTTCAGAGCGGACTCTG 331
Qy 123 LysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSerLeuAla 142
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 332 CAGCCACGACGTTTCTGGAAATATTTCCGATCCCTATACATAGTACACGACCTGTG 391
Qy 143 ThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLysAlaGlu 162
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 392 ACTTCTTCCATTAATACCTCTCCATCATTTCTCTACCCAAATATCTCAAGACTGAA 451
Qy 163 IleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlnAsn 182
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 452 ATAAATGTTCCAGAGTCAAGAGTGAATTTGAGGAGTGAAGTGAAGTGAAGTGAAGT 511
Qy 183 LysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArgValLeu 202
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 512 GAGACCTCCAACTGTGAGGAGTTTAAGAGAGAACAAATGAAGAGAGTGAAGAGAGT 571
Qy 203 CysGlyGluGlnGlnAlaAspAlaAlaGlyAlaGlnValCysSerLeuLeuAla 222
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 572 TGTCCAGGAGAACAG-----GCCAGGTGTCTCTCTGCTCTGCTGCTGCTG 613
Qy 223 GlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGluLeuSer 242
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 614 CAGTCTGAGGTGAAGCCCTCAGTGCCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 673
Qy 243 SerLysLeuGlnLeuMetLysHisGlnSerAspLeuLysLysLeuGlyLeuLeuAsp 262
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 674 AGCAAGATCAGACTTCTGAAAGGTACCAGTGTGACCTGAGAGAGATGGGACAGAC 733
Qy 263 PheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIleAlaLeu 282
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 734 ATCTCTGAAGAAGATGTCCAGCACCAGAGCTATTCCCGGAGACCTTGATTCGACTG 793
Qy 283 ValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMetAsnArg 302
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 794 GTCACCTCGGGGATCCTGCTGGCCATCTGTATCATCACCAGTGTCTATTTCTGTGAACCG 853
Qy 303 ArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrTyrThrGluAsnGly 322
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 854 CGCAGCTCGAGCCCGCAGCAGAGAAAGGCTGGCGGAGACCCCTTATTACGTGGAGAACGG 913
Qy 323 GlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLysAlaSer 342
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 914 GGAGCCAGGCGTATTAATCAGGCTCTGAGGCTCTCCCGAGGCTCAGGGAAGGCCATT 973
Qy 343 ValAsnArgGlyAlaGlnGlnAsnGlyThrGlyGlnAlaThrSerArgAsnGlyHisSer 362
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 974 GTGAATCGAGGGGCTCAGGAGAAATGGCAGCCAGGACACCTCCAGAAACGGCCATTCA 1033
Qy 363 AlaArgGlnHisValValAlaAspThrGluLeu 373
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1034 GCGAGACAAACGTGTGTAGTGTATACCGAAGT 1066

RESULT 12
AF202879
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Homo sapiens hematopoietic progenitor cell antigen CD34 precursor
(CD34) mRNA, partial cds.
AF202879
AF202879.1 GI:6503195
human.
REFERENCE
1 (bases 1 to 581)

```

AUTHORS Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.
 TITLE Direct isolation and cDNA sequencing of mRNA from human
 JOURNAL hematopoietic progenitor cell antigen CD34
 REFERENCE Unpublished
 2 (bases 1 to 581)
 AUTHORS Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.
 TITLE Direct Submission
 JOURNAL Submitted (08-NOV-1999) Medicine, Hematology and Oncology, Sickle
 Cell Center, Medical College of Georgia, 15th Street, AC-1000,
 Augusta, GA 30912, USA

FEATURES
 source Location/Qualifiers
 i. .581
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1q32"
 <1..>581
 /gene="CD34"
 <1..>581
 /gene="CD34"
 /codon_start=2
 /product="hematopoietic progenitor cell antigen CD34
 precursor"
 /protein_id="AAF14634.1"
 /db_xref="GI:6503196"
 /translation="ILSDIKAEIKSGIREVKLTQICLEQNKTSSCAEFKKDRGEGL
 ARVLCGEQADADAGQVCSLLLAQSEVRPOCLLLVANRTEISSKLQMLKKHQSDLK
 KLGLDFTQDVASHQSYOKTLIALVTSGALLAVLGITGYELMNRRSWPTGERLGE
 DPYYTENGSGGQYSSGPGTSPQAQKASVNRGA"

BASE COUNT 150 a 146 c 171 g 114 t

ORIGIN

Alignment Scores:
 Pred. No.: 5,01e-56 Length: 581
 Score: 980.00 Matches: 193
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 51.72% Indels: 0
 DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AF202879 (1-581)

QY 155 IleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThr 174
 |||||
 Db 2 ATCCTAAGTGACATCAAGGCAGAAATCAAAATGTTCAAGGCATCAGAAAGTGAATTTGACT 61
 |||||

QY 175 GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArg 194
 |||||
 Db 62 CAGGGCATCTGCCCTGGAGCAAAATAGACCTCCAGCTGTGCGGAGTTTAAGAAAGGACAGG 121
 |||||

QY 195 GlyGluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAla 214
 |||||
 Db 122 GGAGAGGCGTGGCCGAGTGTGTGTGGGAGGACGAGCTGATGCTGATGTGGGGCC 181
 |||||

QY 215 GlnValCysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuVal 234
 |||||
 Db 182 CAGGTATGCTCCCTGCTCTGCCAGTCTGAGGTGAGGCCTCAGTGTCTACTGCTGGTC 241
 |||||

QY 235 LeuAlaAsnArgThrGluIleSerSerLysLysLeuGlnLeuMetLysLysHisGlnSerAsp 254
 |||||
 Db 242 TTGGCCCAACAGAACAGAAATTTCCAGCAACTCCAACTTATGAAAAGCACCACCAATCTGAC 301
 |||||

QY 255 LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyr 274
 |||||
 Db 302 CTGAAAAGCTGGGATCTCTAGATTCTAGTACAGCAAGATGTTGCAAGCCACCAAGACTAT 361
 |||||

QY 275 SerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIle 294
 |||||
 Db 362 TCCCAAAAGACCCCTGATTGCACCTGGTCACTCGGGAGCCCTGCTGCTCTTGGGCATC 421
 |||||

QY 295 ThrGlyTyrPheLeuMetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGlu 314
 |||||
 Db 422 ACTGGCTATTTCTGTGATGAATCGCCGACAGCTGGAGCCCAACAGAGAAAGGCTGGCGCAA 481
 |||||

QY 315 AspProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSer 334
 |||||
 Db 482 GACCCATTATTACAGGAAACGGTGGAGGCCAGGGCTATTAGCTCAGGACCTGGGACCTCC 541
 |||||

QY 335 ProGluAlaGlnGlyLysAlaSerValAsnArgGlyAla 347
 |||||
 Db 542 CCTGAGGCTCAGGAAAGGCCAGCTGTGAACCGAGGGGCT 580
 |||||

RESULT 13
 AR078454
 LOCUS AR078454 1065 bp DNA linear PAT 31-AUG-2000
 DEFINITION Sequence 18 from patent US 5962644.
 ACCESSION AR078454
 VERSION AR078454.1 GI:10005200
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1065)
 AUTHORS Hawley,R.J. and Monroy,R.L.
 TITLE Porcine CD34
 JOURNAL Patent: US 5962644-A 18 05-OCT-1999;
 FEATURES Location/Qualifiers
 source
 1..1065
 /organism="unknown"
 BASE COUNT 248 a 326 c 273 g 218 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.04e-50 Length: 1065
 Score: 904.00 Matches: 196
 Percent Similarity: 69.44% Conservative: 38
 Best Local Similarity: 58.16% Mismatches: 77
 Query Match: 47.70% Indels: 26
 DB: 6 Gaps: 6

US-09-836-602-2 (1-373) x AR078454 (1-1065)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 |||||
 Db 67 ATGCCCGGGGGTGGACACAGCTCTGCTTGTCTGAGTTGCTGCCCTTGGGTTTCACAGCT 126
 |||||

QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuPro----- 32
 ::: |||::: |||||

Db 127 GTG---AACAGCTCAACTATTGCTTCCACCTTGCCAGCTGCCCTGGGTCACTCCACC 183
 |||||

QY 33 -----ThrGlnGlyThrPheSerAsn----- 39
 |||||

Db 184 GGCCCGGCTACCGCAGGGGCGAGCTATCACCGGGTCAACTATCTCAGACATATCTTCACCT 243
 |||||

QY 40 ValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSer 59
 |||||
 Db 244 GTTCTTACAAATATATCCAAACGAGGAACACACATCA---GATGCTTTCGAAAGTGCCAG 300
 |||||

QY 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVal 79
 |||||

Db 301 CTCACACTGCTCTCTCAGGGCAGCAGTGGGACCACCGTAGCCATCTCAGGCCCTACAGTT 360
 |||||

QY 80 LysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGln 99
 |||||

Db 361 AATTTTCATGCTACCTCGGGGGTCACTCCCGGAAACCGTTAACTCTTCTGTGCCAG 420
 |||||

QY 100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
 |||||

Db 421 CCTCAGACCTCT---CTAGCCACAGCGTCTCTCGGCCACCATCAACTTTCAACTTCAGAG 477
 |||||

QY 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 139
 |||||

Db 478 GTGACCTCGACCCACGACGCTTCCAGGAAATGTTTTCAGACCCCTCTACACACAGTACC 537
 |||||

QY 140 SerLeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIle 159
 |||||

Alignment Scores:					
Pred. No.:	5.64e-30	Length:	160771		
Score:	630.50	Matches:	185		
Percent Similarity:	29.32%	Conservative:	0		
Best Local Similarity:	29.32%	Mismatches:	4		
Query Match:	33.27%	Indels:	445		
DB:	9	Gaps:	3		
US-09-836-602-2 (1-373) x HS88L2 (1-160771)					
DB	26295	GGCTATTTCCTGATGAAATCGCCCGCAGCTGGAGCCCCCACCAGGAGAGAAAGGCTT-GGTCAAGTTC	312	-----	312
Qy	312	-----	-----	-----	-----
DB	26236	TGGGGCCAGGTTAAAGGAAATGAGGAAGATAGTGGGTTTCTGGGAGTTTCAGTGGATGT	26177	-----	26177
Qy	312	-----	-----	-----	-----
DB	26176	CATGGACGAGGAGGAGAAATACTAGAAAAAGCCCTTCTGTGAGCTTACATAAAGATATGC	26117	-----	26117
Qy	312	-----	-----	-----	-----

```
Db 26116 ATGTGTGCACACACAGTAATCGTGGAGATTCAAAATATATATGTAGCAAACTGGGGAGG 26057
QY 312 ----- 312
Db 26056 GGACAGTAATGGTGGCAGCCACCTACTCGGTGTGTAGCAGATGATTCGTGTCACTCTAT 25997
QY 312 ----- 312
Db 25996 TCATTATGTTTTCTGGATAGTTTCAGCTTGGGGCCACCAGAACAGGCTGTTTAATCAGCC 25937
QY 312 ----- 312
Db 25936 ACACCTGATTTGCAACCGTGTAAATGCCAGTGGGTGCCCTTGCTCAAAAGGAGGTA 25877
QY 312 ----- 312
Db 25876 TACGGAGGAGAAATCCCATTTGCCATTCTGGATGAGGAGGACAAAGGTGAGGTCTGAA 25817
QY 312 ----- 312
Db 25816 TCTTGGCCTCTGGCCTGTCCCTTACCCTGGGAGGTTCATCCACCCCTTCTTGAAGCTGCC 25757
QY 312 ----- 312
Db 25756 CGTTTTCCCGGGAGCTGGCTGCCGAGCTGCCGTGGCATGGTGTGCCCTCTCCTATG 25697
QY 312 ----- 312
Db 25696 TCCTTTCTCCTTAGGAGCTGGAACCTTGACCCTTTCAGGAAGAAAGAGTCTGCACA 25637
QY 312 ----- 312
Db 25636 TGCAGCTGCACCCCTCCCTCCGATCCTTCTCCACCTCCCTCCCTTCTCCACCCC 25577
QY 312 ----- 312
Db 25576 TGCCCCCACTTCTGTGTTTGGGCCCTCTCCCATCCAGTGTCTCACAGCCCTGCTTACCAG 25517
QY 313 -----GlyGluAspProTyrThrGluAsnGly 322
Db 25516 ATAACTCTACTTTATTTATACACTGTCTAGGGCGAAGACCCCTATTATACACGGAACCGT 25457
QY 323 GlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLysAlaSer 342
Db 25456 GGAGGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTGAGGCTCAGGGAAGGCCAGT 25397
QY 343 ValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGlyHisSer 362
Db 25396 GTCAACCGAGGGGCTCAGGAAACGGGACCGGCCAGCCCTCCAGAACCGCCATTCA 25337
QY 363 AlaArgGlnHisValValAlaAspThrGluLeu 373
Db 25336 GCAAGACAAACACGTGGTGGCTGATACCGAATTG 25304
```

Search completed: October 30, 2002, 09:51:15
Job time : 1971.1 secs